

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Departamento de Biologia Geral
Programa de Pós-Graduação em Genética

Hatilla dos Santos Silva

**ANCESTRALIDADE GENÉTICA E PERFIL DE PRODUÇÃO DE CITOCINAS
EM UMA POPULAÇÃO DE CRIANÇAS BRASILEIRAS**

Belo Horizonte
2022

Hatilla dos Santos Silva

**ANCESTRALIDADE GENÉTICA E PERFIL DE PRODUÇÃO DE CITOCINAS
EM UMA POPULAÇÃO DE CRIANÇAS BRASILEIRAS**

Versão final

Dissertação apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Minas Gerais como pré-requisito parcial para a obtenção do título de Mestre em Genética.

Orientador: Eduardo Martin Tarazona-Santos

Coorientadora: Camila Alexandrina Viana Figueiredo

Belo Horizonte
2022

043

Silva, Hatilla dos Santos.

Ancestralidade genética e perfil de produção de citocinas em uma população de crianças brasileiras [manuscrito] / Hatilla dos Santos Silva. – 2022.

70 f. : il. ; 29,5 cm.

Orientador: Eduardo Martin Tarazona-Santos. Coorientadora: Camila Alexandrina Viana Figueiredo.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Genética. 2. Imunogenética. 3. Células Th2. 4. Citocinas. 5. Interleucina-13. I. Tarazona-Santos, Eduardo Martin. II. Figueiredo, Camila Alexandrina Viana. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 575



UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Genética

ATA DE DEFESA DE DISSERTAÇÃO

ATA DA DEFESA DE DISSERTAÇÃO	326/2022 entrada
Hatilla dos Santos Silva	1º/2020 CPF: 860.371.065-17

Às oito horas e trinta minutos do dia **05 de julho de 2022**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**Ancestralidade genética e perfil de produção de citocinas em uma população de crianças brasileiras**", requisito para obtenção do grau de Mestre em **Genética**. Abrindo a sessão, o Presidente da Comissão, **Eduardo Martin Tarazona Santos- Orientador**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Eduardo Martin Tarazona Santos-Orientador	UFMG	012.494.056-02	APROVADO
Camila Viana Figueiredo	UFBA	007.601.734-60	APROVADO
Alessandra Pontillo	USP	234.006.268-30	APROVADO
Marcelo Rizatti Luizon	UFMG	277.308.188-92	APROVADO

Pelas indicações, o candidato foi considerado: APROVADO

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 05 de julho de 2022.

Eduardo Martin Tarazona Santos- Orientador

Camila Viana Figueiredo

Alessandra Pontillo

Marcelo Rizatti Luizon



Documento assinado eletronicamente por **Alessandra Pontillo, Usuário Externo**, em 05/07/2022, às 10:38, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Camila Alexandrina Viana de Figueiredo Fontana, Usuário Externo**, em 05/07/2022, às 10:56, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Marcelo Rizzatti Luizon, Professor do Magistério Superior**, em 05/07/2022, às 13:34, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eduardo Martin Tarazona Santos, Professor do Magistério Superior**, em 05/07/2022, às 14:23, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1567455** e o código CRC **F3D41474**.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Genética

FOLHA DE APROVAÇÃO

"Ancestralidade genética e perfil de produção de citocinas em uma população de crianças brasileiras"

Hatilla dos Santos Silva

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Eduardo Martin Tarazona Santos - Orientador
UFMG

Camila Viana Figueiredo
UFBA

Alessandra Pontillo
USP

Marcelo Rizatti Luizon
UFMG

Belo Horizonte, 05 de julho de 2022.



Documento assinado eletronicamente por **Alessandra Pontillo, Usuário Externo**, em 05/07/2022, às 10:38, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Camila Alexandrina Viana de Figueiredo Fontana, Usuário Externo**, em 05/07/2022, às 10:56, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Marcelo Rizzatti Luizon, Professor do Magistério Superior**, em 05/07/2022, às 13:34, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eduardo Martin Tarazona Santos, Professor do Magistério Superior**, em 05/07/2022, às 14:24, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1567533** e o código CRC **9BDD04EF**.

AGRADECIMENTOS

À Universidade Federal de Minas Gerais e à Pós-Graduação em Genética, por ter me propiciado diversos aprendizados e oportunidades durante o mestrado, colaborando assim para meu crescimento acadêmico e pessoal. À FAPEMIG pelo suporte financeiro, apoiando e permitindo minha qualificação acadêmica e profissional. Aos projetos EPIGEN e SCAALA, assim como aos colaboradores envolvidos, em especial ao Dr. Maurício Barreto, e a todos os participantes sem os quais este trabalho não seria possível. Aos meus orientadores Eduardo Tarazona Santos e Camila Alexandrina Figueiredo, pela confiança e oportunidade, me orientando com muita dedicação, viabilizando a realização do mestrado. Eduardo e Camila são líderes admiráveis, com muito comprometimento com a produção científica e formação dos seus estudantes. Agradeço o Laboratório de Diversidade Genética Humana (LDGH - UFMG) e o Laboratório de Imunofarmacologia e Biologia Molecular (IMUNOBIO - UFBA), e todos seus integrantes, por possibilitar a realização do projeto, além de aprender novos conhecimentos e ferramentas da Genética Humana. Agradeço especialmente a Camila Zolini de Sá (LDGH), por toda ajuda em diversas situações. Agradeço imensamente Helena Mariana Pitangueira (IMUNOBIO) por todo auxílio e ensinamentos que foram cruciais para a execução desse projeto. Às professoras e professores que tive a honra de conhecer ao longo da graduação e mestrado, por todos os ensinamentos, em especial ao professor Ryan dos Santos Costa por ter me oportunizado o envolvimento com a pesquisa e consequente meu crescimento científico. A todos da família que de alguma forma contribuíram na minha formação, seja pelo amor e carinho, pelos incentivos, pela torcida e apoio nas mais diversas situações na minha vida. Agradeço especialmente à minha mãe, Ana Cristina, por sempre ter sido presente ao longo da minha caminhada, me ajudando de incontáveis formas, mas especialmente me dando muito amor, carinho e incentivo. Ao meu companheiro, Juan Vitor, agradeço por todo o incentivo, suporte e contribuição nas mais diversas situações, especialmente me ajudando a sistematizar e organizar as coisas. Aos meus amigos e amigas, agradeço por sempre terem dado força para encarar novos desafios. Em especial a Adson, Ana Paula, Antonio Mateus, Jeferson, Juliana, Maria, Raísa, Stefane, Talita, Victor e Vida, que estão ao meu lado tornando a vida acadêmica mais leve. Agradeço a todos que, mesmo não sendo citados, colaboraram para a construção e conclusão dessa importante etapa.

Resumo

O perfil de citocinas produzido pelas células CD4+ é um marcador importante do tipo de resposta imune, por exemplo, as respostas Th1 (IFN- γ), Th2 (Interleucina (IL)-5 e IL-13) e Treg (IL-10). Ademais, a resposta imune é considerada um traço complexo resultante da interação entre fatores ambientais, genéticos e epigenéticos. Nesta perspectiva, a base genética dos componentes imunológicos é produto de processos evolutivos que atuaram ao longo da história humana, especialmente considerando a diversidade de ambientes encarada ao longo da dispersão da espécie. Assim, o presente trabalho tem como objetivo avaliar a influência da ancestralidade genômica na produção das citocinas IFN- γ (Th1), IL-5 (Th2), IL-13 (Th2) e IL-10 (Treg) em uma população brasileira. Os participantes do estudo são crianças (4 -11 anos) que fazem parte do programa SCAALA (*Social Change Asthma and Allergy in Latin America*) em Salvador - Bahia. Foi realizada a estimativa de ancestralidade global e local a partir de variantes genotipadas utilizando o chip Illumina Omni 2,5M. A produção das citocinas foi mensurada a partir da cultura de células sanguíneas por meio do imunoenensaio ELISA. A associação entre ancestralidade global e produção de citocinas foi investigada por meio de regressão logística, incluindo variáveis potencialmente confundidoras. Foi realizado o mapeamento por miscigenação para a citocina IL-13 utilizando o software PLINK. Encontramos uma associação entre ancestralidade genômica biogeográfica e produção da citocina da assinatura Th2, IL-13, em que indivíduos com maior ancestralidade africana tiveram uma menor chance de serem produtores de IL-13. A análise posterior de mapeamento por miscigenação permitiu a identificação de três variantes intrônicas (rs12598402-*PLCG2*, rs28376615-*CDH13* e rs75929636-*CDH13*) associadas com a produção de IL-13. Além disso, os alelos destas três variantes que estão associados a uma maior produção de IL-13 têm uma maior frequência em indivíduos com maior influência genética africana em seu genoma. Estudos futuros são necessários para melhor elucidação dos mecanismos associados à regulação da resposta Th2 na população Brasileira.

Palavras-chave: Imunogenética, Th2, IL-13, Citocinas, Ancestralidade

Abstract

The profile of cytokines produced by CD4⁺ cells is an important marker of the type of immune response, for example, Th1 (IFN- γ), Th2 (Interleukin (IL)-5 and IL-13), and Treg (IL-10) responses. Furthermore, the immune response is considered a trait resulting from the interaction between environmental, genetic, and epigenetic factors. In this perspective, the genetic basis of the immunological components is a product of evolutionary processes that acted throughout human history, especially considering the diversity of environments faced throughout the dispersion of species. Thus, the present study aims to evaluate the influence of genomic ancestry on the production of IFN- γ (Th1), IL-5 (Th2), IL-13 (Th2), and IL-10 (Treg) cytokines in a Brazilian population. Study participants are children (4-11 years old) who are part of the SCAALA program (Social Change Asthma and Allergy in Latin America) in Salvador - Bahia. Global and local ancestry estimation was performed from genotyped variants using the Illumina Omni 2.5M chip. The production of cytokines was measured from the culture of blood cells through the immunoassay ELISA. The association between global ancestry and the production of cytokines was investigated using logistic regression, including potential confound variables. Admixture mapping for the IL-13 cytokine was performed using PLINK software. Here, we found an association between biogeographic genomic ancestry and production of the Th2 signature cytokine IL-13, in which individuals with greater African ancestry had a lower chance of producing IL3. A fine-mapping analysis reveals three variants (rs12598402-*PLCG2* and rs28376615-*CDH13*, rs75929636-*CDH132*) associated with the production of IL-13. In addition, alleles of these three variants that were associated with a higher production of IL-13 were more frequent in individuals with greater African genetic influence in the genome. Future studies are needed to better elucidate the mechanisms associated with the regulation of Th2 response in the Brazilian population.

Keywords: Immunogenetics, Th2, IL-13, Cytokines, Ancestry

LISTA DE FIGURAS

Figura 1 – Diferenciação dos subtipos de células T <i>helper</i>	15
Figura 2 – Fatores determinantes dos diversos fenótipos imunológicos.....	17
Figura 3 – Hipótese da Biodiversidade/Velhos amigos e o Sistema Imune.....	24
Figura 4 – Principais rotas migratórias seguidas pelos ancestrais humanos que se originaram na África.....	25
Figura 5 – Diferença entre ancestralidade genética global (A) e local (B) em uma população iscigenada.....	27
Figure 1 – Manhattan plots of the admixture mapping (AM) analyses for IL-13 production.....	39
Figure 2 – Frequency distribution of genotypes according to the estimated African genomic ancestry for the SCAALA population.....	41
Figure S1 – Global ancestry quantiles comparison for IL-13 concentration.....	57
Figure S2 – Quantile-quantile (QQ) plot for IL-13 production association with African Local Ancestry.	57
Figure S3 – Quantile-quantile (QQ) plot for IL-13 production association with European Local Ancestry.	58
Figure S4 – Quantile-quantile (QQ) plot for IL-13 production association with Native American Local Ancestry.	58
Figure S5 – Linkage Disequilibrium (LD) Plot for associated variants rs12598402, rs28376615 and rs75929636.....	59
Figure S6 – IL-13 concentration and rs12598402 (A), rs28376615 (B) and rs75929636 (C) genotype.	59
Figure S7 – Violin plot of <i>PLCG2</i> intron-excision ratios according to rs12598402 genotypes in Genotype-Tissue Expression (GTEx)	60
Figure S8 – Frequency distribution of genotypes according to the estimated European genomic ancestry for the SCAALA population.	60

LISTA DE TABELAS

Table 1 – Association of African ancestry proportion and quantiles with cytokines production.	37
Table 2 – Admixture mapping analysis suggestive peaks for IL-13 production among children from the Social Change, Asthma, Allergy in Latin America (SCAALA) Cohort in Salvador, Brazil.	40
Table 3 – Fine-mapping significant associations for IL-13 production identified through imputed variants.	40
Table S1 – Populations used for ancestry inference according to continental origin and data source.	50
Table S2 – IFN- γ production in SCAALA cohort according to sociodemographic and environmental variables.	51
Table S3 – IL-5 production in SCAALA cohort according to sociodemographic and environmental variables.	52
Table S4 – IL-13 production in SCAALA cohort according to sociodemographic and environmental variables.	53
Table S5 – IL-10 production in SCAALA cohort according to sociodemographic and environmental variables.	54
Table S6 – Admixture Mapping statistics of identified suggestive peaks for IL-13 production among children from the Social Change, Asthma, Allergy in Latin America (SCAALA) Cohort in Salvador, Brazil.	55
Table S7 – Local ancestry window in 16q23.3 associations for IL-13 adjusted for previously associated variants.	56

LISTA DE ABREVIATURAS E SIGLAS

AFR	Africanos
APC	<i>Antigen Presenting Cell</i>
CCL	<i>C-C motif chemokine ligand</i>
CDH13	<i>Cadherin 13</i>
CNV	<i>Copy Number Variation</i>
ELISA	<i>Enzyme-linked immunosorbent assay</i>
EUR	Europeus
GWAS	<i>Genome-Wide Association Study</i>
HFGP	<i>Human Functional Genomics Project</i>
IFN	Interferon
IL	Interleucina
NAT	Nativos Americanos
NK	<i>Natural Killer Cell</i>
OOA	<i>Out of Africa</i>
PBMC	<i>Peripheral Blood Mononuclear Cell</i>
PLCG2	<i>Phospholipase C gamma 2</i>
QTL	<i>Quantitative Trait Locus</i>
RI	Resposta Imune
SCAALA	<i>Social Changes, Asthma and Allergy in Latin America</i>
SI	Sistema Imune
SNV	<i>Single Nucleotide Variant</i>
TCR	<i>T-cell Receptor</i>
Th	<i>T-helper</i>
Treg	T-regulatória

SUMÁRIO

1. INTRODUÇÃO	13
1.1. Visão geral da resposta imune.....	13
1.2. O Sistema Imune e suas variações.....	16
1.3. Fatores genéticos.....	18
1.4. Fatores não-genéticos.....	20
1.5. Origens da variação interpopulacional na RI.....	21
1.6. Evolução humana e a arquitetura genética da resposta imune.....	23
1.7. Ancestralidade genética, miscigenação e produção de citocinas.....	26
1.8. Justificativa.....	28
2. CAPÍTULO ÚNICO	29
Abstract.....	29
Introduction.....	30
Methods.....	32
Results.....	36
Discussion.....	41
References.....	45
Supplementary Material.....	50
3. DISCUSSÃO	61
4. CONCLUSÃO	63
REFERÊNCIAS	64

1. INTRODUÇÃO

1.1. Visão geral da resposta imune

Os humanos, assim como outros seres, estão constantemente expostos a organismos patogênicos e não-patogênicos, que por sua vez podem conter inúmeras moléculas que impactam o funcionamento do organismo. Neste contexto, o Sistema Imunológico (SI) tem um papel essencial na manutenção e recuperação da homeostase. Isso se dá diante de respostas imunológicas montadas contra agentes externos, mas também contra internos, funcionando por meio de mediadores celulares e moleculares que constituem a resposta imune (RI) (BRODIN; DAVIS, 2017). A RI do hospedeiro pode ser conceitualmente classificada em inata, cujo os mecanismos envolvem o reconhecimento de padrões moleculares associados aos patógenos, e adaptativa, que é capaz de reconhecer e montar uma resposta específica. Vale ressaltar que esses tipos de respostas não atuam de maneira independente, ao contrário, são mecanismos complementares e correlacionados, aspectos fundamentais para o funcionamento correto do SI (MARSHALL *et al.*, 2018). Por exemplo, o desenvolvimento da resposta imune adaptativa é dependente da apresentação do antígeno por células da resposta imune inata.

Um dos principais mecanismos moleculares associados à resposta imune é a produção de citocinas, tanto por células da RI inata quanto adaptativa, cujo o papel é essencial na montagem e regulação da resposta imune (KANY *et al.*, 2019; LIU *et al.*, 2021). As citocinas são um grande grupo de proteínas de baixo peso molecular envolvidas na sinalização celular no contexto da RI, são principalmente produzidas por leucócitos, mas outros tipos celulares também podem produzir tais sinalizadores (KANY *et al.*, 2019). Além disso, populações celulares distintas podem produzir uma mesma citocina. A atividade efetora destas moléculas se dá por interação com receptores celulares, que após ativados resultam em uma cascata de sinalização celular, levando a ativação ou inativação de células específicas, liberação de outros mediadores moleculares, entre outras ações (KANY *et al.*, 2019). Uma característica chave das citocinas é a sua ação pleiotrópica em que os efeitos provocados por essas moléculas sinalizadoras dependem da célula que recebe o sinal.

Em condições adequadas, as células T podem ser estimuladas por células apresentadoras de antígenos (APCs), levando a diferenciação em dois subtipos

primários, as células T citotóxicas (CD8+) e as células *T-helper* (CD4+) (IWASAKI *et al.*, 2015). As células Th (T Helper) têm um papel importante no estabelecimento e amplificação da resposta imune. Tais células, uma vez ativadas por meio da interação com seus receptores de membrana, os TCRs, tem sua ação principalmente mediada por citocinas que modulam e influenciam a atividade de diversas outras células (LUCKHEERAM *et al.*, 2012).

As células Th podem ser diferenciadas em subpopulações de acordo com o microambiente ao qual estão expostas, resultando em polarização em perfis distintos de produção de citocinas. Dados os estímulos adequados, ocorre em células T *naive* a ativação de fatores de transcrição envolvidos na síntese de diversos receptores e mediadores, incluindo as citocinas, característicos da respectiva resposta (PAWLAK *et al.*, 2020). Dentre os vários perfis existentes, os principais e mais estudados são os tipos Treg, Th1, Th2 e Th17 (LUCKHEERAM *et al.*, 2012; EIZENBERG-MAGAR, 2017; MARTINEZ-SANCHEZ *et al.*, 2018; ZHU *et al.*, 2019). Ademais, descobertas mais recentes têm mostrado que as subpopulações de células Th não estão em uma diferenciação final, como se pensava anteriormente, ao contrário, têm se observado o fenômeno de plasticidade onde células T polarizadas mudam seu fenótipo para um misto (*e.g.*, Th1 → Th1/Th17) ou alternativo (*e.g.*, Th1 → Th17) (DUPAGE *et al.*, 2016). A Figura 1 apresenta os principais subtipos de células Th e o perfil de citocinas correspondente a cada fenótipo.

A resposta do tipo Th1 está principalmente relacionada à produção da citocina IFN- γ , que por sua vez ativa a atividade bactericida de macrófagos bem como aumenta a imunidade contra patógenos intracelulares (MARSHALL *et al.*, 2018). Falhas nos mecanismos de montagem da resposta Th1 de maneira adequada estão implicadas com o desenvolvimento de certas doenças autoimunes como Diabetes tipo 1, Esclerose Múltipla e Artrite Reumatoide (RAPHAEL *et al.*, 2015).

Em contrapartida, a resposta do tipo Th2 tem como citocinas assinatura as Interleucinas (IL)-4, IL-5 e IL-13, por exemplo, e está envolvida no desenvolvimento de células B produtoras do anticorpo IgE (Imunoglobulina E). Além disso, também está envolvida no recrutamento de mastócitos e eosinófilos, células efetoras essenciais na montagem da RI contra diversos parasitas multicelulares (MARSHALL *et al.*, 2018). Esse perfil celular característico da resposta Th2 também está associado à alergia e

alguns fenótipos de asma, sendo observado na resposta inflamatória dessas condições. Acredita-se, dessa forma, que o desbalanço na produção de citocinas Th2 está associado com o desenvolvimento de atopia (alergia), por exemplo.

As células Th17 foram descritas mais recentemente, caracterizadas pela produção de citocinas da família IL-17, estão associadas com diversas condições inflamatórias, tanto em infecções crônicas quanto em algumas doenças. Por exemplo, o perfil Th17 está envolvido na patogênese tanto de doenças autoimunes quanto em alguns subtipos de asma (BEDOYA *et al.*, 2013; ZHAO *et al.*, 2013).

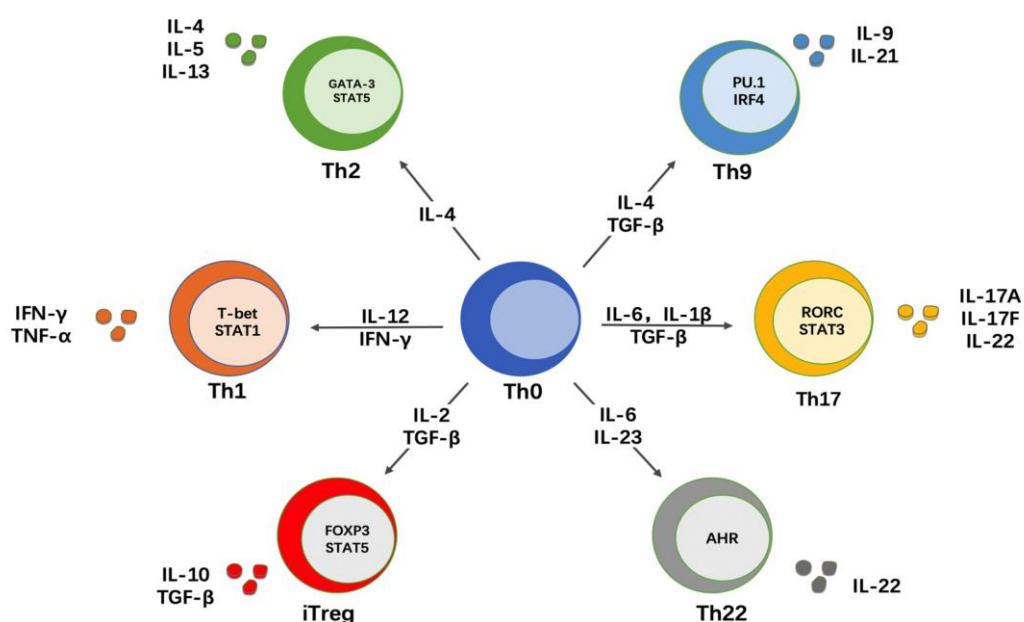


Figura 1. Diferenciação dos subtipos de células T helper. Células T naïve podem se diferenciar em subpopulações distintas dependendo do contexto de citocinas ao qual estão expostas. Estas subpopulações têm sua diferenciação mediada pela ativação de fatores de transcrição, levando à expressão de citocinas de assinatura relacionadas à função efetora de cada subconjunto. Lee *et al.*, 2021

Uma outra subpopulação de células CD4⁺ são conhecidas como células regulatórias (Treg) (BUSZKO *et al.*, 2020) cuja ação está ligada à produção de citocinas que podem exercer uma ação regulatória em diferentes contextos, como a IL-10. O papel central desse grupo de células é modular a resposta imune de diferentes formas, muitas vezes suprimindo a resposta inflamatória e regulando negativamente a produção de outras citocinas. Além disso, outros perfis vêm sendo descritos e caracterizados como o Th3, Th9, Th22, Tfh, entre outros, o que traz ainda mais complexidade na compreensão dos mecanismos associados às células CD4⁺ (RAPHAEL *et al.*, 2015).

1.2. O Sistema Imune e suas variações

Ao estudar a variação do perfil imunológico individual é possível notar que após a exposição a algum fator que afeta a homeostase, tal como uma infecção aguda ou o processo de vacinação, há uma mudança no *status* da resposta imune. Entretanto, após a resolução, o sistema imune tende a retornar ao estado em que se encontrava previamente à exposição, em homeostase (BRODIN; DAVIS, 2017). Tal ideia vem sendo corroborada por estudos longitudinais, em que há mensuração dos parâmetros imunológicos em indivíduos ao longo do tempo. Por exemplo, amostras de sangue de indivíduos adultos saudáveis coletadas ao longo de semanas ou meses mostram que há uma estabilidade na frequência de células e também nos níveis séricos de proteínas (TSANG *et al.*, 2014; SHEN-ORR *et al.*, 2016; CARR *et al.*, 2016). Tais estudos demonstram que cada organismo possui um status basal na composição do sistema imune, que tende a se manter estável ao longo do tempo. Isso não significa que tal status não possa ser alterado, ou que permanece de uma mesma forma ao longo de toda a vida, especialmente considerando que a composição do sistema imune em um mesmo indivíduo pode ser distinta em diferentes faixas etárias (SIMON *et al.*, 2015).

Diversos estudos envolvendo a investigação de parâmetros imunológicos em uma perspectiva populacional corroboram que os sistemas imunológicos humanos são altamente variáveis entre os indivíduos, e até mesmo, entre populações (KOLLMANN *et al.*, 2013; BRODIN *et al.*, 2015; SANZ *et al.*, 2018). Essa variação pode se manifestar de distintas formas, tanto a nível celular quanto molecular, por exemplo, diferenças na frequência de populações e subpopulações de leucócitos, ou até mesmo na expressão e produção de fatores de transcrição, receptores e outras moléculas (TSANG *et al.*, 2015; LISTON *et al.*, 2016). Brodin e Davis (2017) comparam a frequência de diversos leucócitos em uma população de indivíduos saudáveis e demonstraram uma grande amplitude na variação em populações específicas de células, como monócitos, células NK e células T. Desta forma, pode-se dizer que a imunidade é um traço intrinsecamente variado entre indivíduos, especialmente considerando as análises realizadas em pessoas aparentemente saudáveis.

Nessa perspectiva, um dos objetivos centrais da imunologia é compreender a variação normal em diversos parâmetros, colaborando para a elucidação dos mecanismos subjacentes bem como os fatores determinantes dos diversos imunofenótipos. Entender tais causas torna-se essencialmente importante quando se considera que certos fenótipos imunológicos podem predispor o desenvolvimento de doenças específicas (BRODIN; DAVIS, 2017). A perspectiva que melhor explica a variação no sistema imune dentro e entre populações inclui fatores ambientais, genéticos e epigenéticos (BARREIRO; QUINTANA-MURCI, 2020). Desta forma, pode-se compreender a expressão do sistema imune individual como um traço geneticamente complexo, produto de diversos fatores e da interação entre eles. A Figura 2 ilustra os fatores que podem influenciar o SI, demonstrado tanto fatores intrínsecos, como extrínsecos ao indivíduo. A contribuição proporcional de fatores genéticos e não-genéticos é alvo de debate, apesar de evidências apontarem que fatores não-genéticos provavelmente têm uma maior contribuição (ORRÚ *et al.*, 2013; BRODIN *et al.*, 2015; BRODIN; DAVIS, 2017; MANGINO *et al.*, 2017). Vale ressaltar que, considerando que o SI é composto por diversas moléculas, células e órgãos, a proporção com que fatores não-genéticos e genéticos contribuem para a expressão de traços imunes depende da unidade alvo em questão.

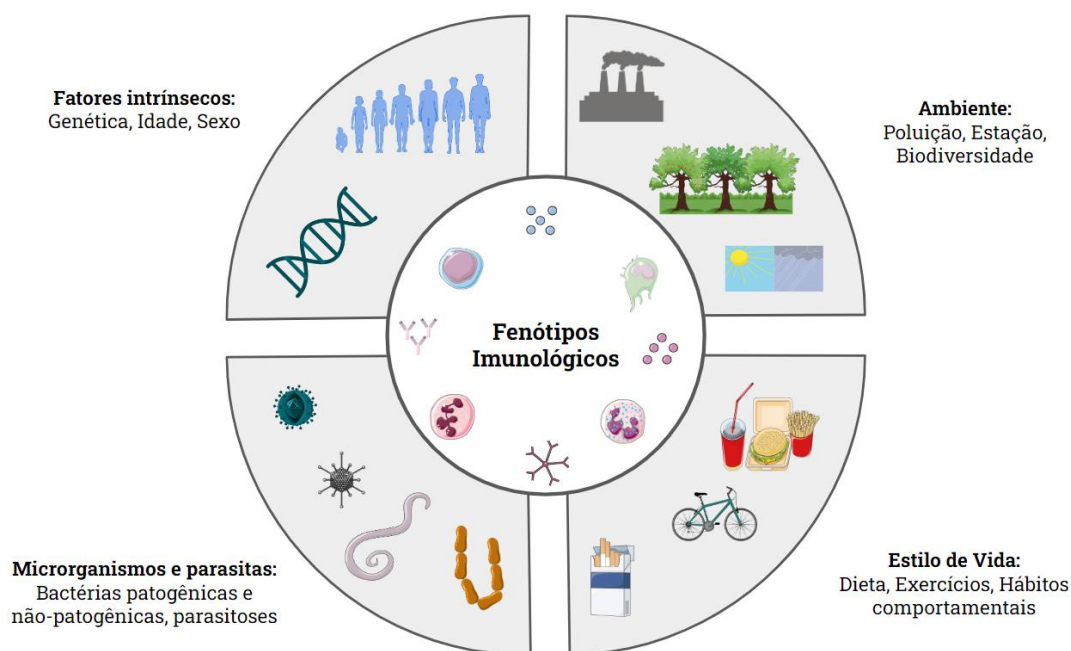


Figura 2. Fatores determinantes dos diversos fenótipos imunológicos. Diferentes fatores que contribuem para variação populacional na imunidade.

1.3. Fatores genéticos

Ainda que as influências ambientais sejam preponderantes na variação interindividual observada nas configurações do sistema imunológico, a genética também apresenta um papel considerável na determinação do status imune. De fato, grandes estudos sobre a base genética da variação nas células e moléculas que compõem o SI explicam em média de 20-40% da variação imunológica total (ORRÚ *et al.*, 2013; BRODIN *et al.*, 2015; MANGINO *et al.*, 2017). Conforme já mencionado, a herdabilidade em parâmetros imunológicos não se distribui uniformemente, logo, o papel da genética dependerá essencialmente do parâmetro analisado, como o tipo celular ou a molécula em questão. Vale ressaltar que há uma crescente influência não-genética com o aumento da idade, isso quer dizer que parâmetros imunológicos podem ser muito mais hereditários se medidos em populações de crianças (BRODIN *et al.*, 2015).

Estudos com gêmeos oferecem uma oportunidade ideal para quantificar a contribuição relativa de fatores hereditários para características imunológicas em humanos saudáveis, permitindo o ajuste para fatores de confusão conhecidos e desconhecidos (VAN DONGEN *et al.*, 2012; SAHU *et al.*, 2016; CRAIG *et al.*, 2020). Nesse sentido, Roeder e colaboradores (2016) estudaram 78.000 traços imunes obtidos por imunofenotipagem em 669 gêmeas, tanto frequência de leucócitos e subpopulações quanto expressão de moléculas de superfície. Os autores encontraram que as estimativas de herdabilidades variaram amplamente, indo de 0%, onde há influências puramente ambientais ou estocásticas, até 96%, indicando um forte efeito genético (ROEDER *et al.*, 2016). Ao investigar as 151 principais características hereditárias, 11 *loci* genéticos foram associados, explicando até 36% da variação de 19 características. Um outro estudo, também com gêmeos, mostrou que 39% das populações de células estudadas apresentaram uma herdabilidade maior que 20%, o limite de detecção do estudo (BRODIN *et al.*, 2015). Apesar dos autores enfatizarem o papel dominante dos fatores ambientais, é notável que a genética também possui alguma contribuição. Por outro lado, Orrú e colaboradores (2013) utilizaram uma outra abordagem por meio do coeficiente de parentesco entre indivíduos (não gêmeos) em que estimaram em 41% a herdabilidade média para diversos tipos e subtipos celulares.

No contexto das citocinas e quimiocinas, foi mostrado que a herdabilidade média foi maior que a encontrada para frequência de células imunes. Os níveis de algumas citocinas se mostraram altamente herdáveis, como para a IL-12p40, enquanto outras, como a IL-10 tiveram uma influência genética menor (BRODIN *et al.*, 2015). A maior influência da genética na produção citocinas e quimiocinas pode estar relacionado ao fato de que a frequência de células é um traço cuja regulação é mais complexa e envolve influências poligênicas, enquanto de quimiocinas e citocinas são produtos diretos da expressão de um único gene (BRODIN *et al.*, 2017). Ademais, os autores avaliaram a resposta celular às citocinas homeostáticas (IL-2 e IL-7) *in vitro* e observaram que a ativação de vias *downstream*, como a fosforilação do fator de transcrição STAT5 foi altamente herdável (BRODIN *et al.*, 2015).

O Projeto Genômica Funcional Humana (Human Functional Genomics Project - HFGP) é uma iniciativa com o intuito de identificar os fatores que explicam a variabilidade da RI, tanto no contexto da saúde quanto de doenças (NETEA *et al.*, 2016). Em um dos estudos realizado pelo grupo, foram investigadas variantes genéticas putativamente associadas com a produção de citocinas em resposta a diversos estímulos (LI *et al.*, 2016).

Em contraste com o estudo de Brodin e colaboradores (2015), eles encontraram que a herdabilidade para a produção estimulada das citocinas em PBMC foi maior que o reportado anteriormente. Ademais, citocinas derivadas de monócitos tiveram uma maior influência de fatores genéticos que as citocinas derivadas de células T. Além disso, o impacto dos fatores genéticos foi diferente entre os estímulos de microrganismos analisados, demonstrando que variantes podem ter um efeito regulatório apenas para alguns patógenos (LI *et al.*, 2016).

Tendo em vista que o background genético pode ter uma influência moderada para alguns traços imunes, um passo importante é a identificação dos *loci* envolvidos na variação fenotípica. Apesar das diversas categorias de variantes genéticas, como as CNVs (variação no número de cópias) ou INDELS (inserção/deleção) potencialmente influenciarem na expressão de traços imunes, as variantes de único nucleotídeo (SNVs) constituem o tipo mais comum de variação ao longo do genoma e são, conseqüentemente, as mais estudadas (LISTON *et al.*, 2013). Além da identificação dos *loci* genéticos, faz-se relevante estabelecer a relações funcionais

que explicam tais associações do ponto de vista molecular. Assim, o mapeamento de *loci* de traços quantitativos (QTL) constitui uma das principais estratégias para compreender a relação genótipo-traço imune, em que diferentes aspectos podem ser estudados como medidas de expressão gênica (eQTLs), níveis séricos de marcadores, como citocinas (cQTLs), ou mesmo a contagem de células (ccQTLs).

Dentre as principais estratégias utilizadas, os estudos de GWAS se destacam, em que, numa perspectiva livre de hipóteses, são investigadas variantes comuns ao longo do genoma. Ademais, a maioria dos QTL identificados para citocinas são do tipo *trans*, ou seja, as variantes estão em genes cuja posição é mais distal em relação ao gene da citocina investigada, sugerindo mecanismos de regulação indiretos (LI *et al.*, 2016). A importância destas análises se torna ainda mais relevante ao se considerar que dados recentes sugerem que eQTLs de resposta imunológica desempenham um papel importante na arquitetura genética de doenças humanas (ROEDER *et al.*, 2015). Li e colaboradores (2016) investigaram o impacto do componente genético na produção estimulada de citocinas e identificaram sinais significativos de variantes associadas (QTLs) ao longo do genoma. Adicionalmente, as variantes identificadas estão mais frequentemente localizadas em regiões sob seleção positiva.

1.4. Fatores não-genéticos

O SI pode ser encarado como sensor de diversos estímulos internos e externos, desta forma, é capaz de modular e ajustar-se de acordo com aquilo que é percebido, tal como alterações ambientais (BRODIN; DAVIS, 2017). As influências não hereditárias têm um papel relevante no desenvolvimento e funcionamento do sistema imune, tais fatores incluem exposições ambientais, exposições a microrganismos patogênicos e simbióticos, bem como fatores intrínsecos do indivíduo como sexo/gênero e idade (BRODIN *et al.*, 2015).

A idade e o sexo são dois fatores que contribuem para a variação no SI. Tem sido demonstrado que o processo de envelhecimento, por exemplo, contribui em média com 5% da variação total (TER HORST *et al.*, 2016). Uma abordagem que também contribui para explicar parcialmente o papel da idade são os estudos com gêmeos, foi observado, por exemplo, que parâmetros imunológicos divergem em

gêmeos monozigóticos conforme a idade aumenta. Nesse sentido, é sugerido que o acúmulo de exposições ambientais contribui para o aumento dessa heterogeneidade. Vale ressaltar que o progresso do sistema imune ao longo dos anos pode ser modulado por fatores ambientais e genéticos. Por exemplo, genes associados com longevidade estão relacionados a mudanças funcionais no SI (BRODIN *et al.*, 2015).

O sexo também se mostra um importante fator no contexto imune, apesar de que são observadas mais diferenças na susceptibilidade a doenças imunomediadas (*e.g.*, Doenças autoimunes) do que em parâmetros imunológicos. A frequências de alguns tipos de linfócitos e produção de citocinas podem variar de acordo com o sexo, tal heterogeneidade pode ser explicada considerando fatores hormonais e fisiológicos distintos (TER HORST *et al.*, 2016). Um ponto crucial nos estudos de variação imune associada ao sexo que é pouco explorado é a discriminação entre o que de fato é atribuído a uma natureza biológica do que é produto de diferenças de gênero, um aspecto mais sociocultural (LISTON *et al.*, 2021). Por exemplo, em um estudo envolvendo análise transcriptômica de células sanguíneas o efeito da variável sexo foi discrepante considerando o ambiente urbano e o ambiente rural em Marrocos (IDAGHDOUR *et al.*, 2010).

Um estudo realizado por Liston e colaboradores (2016) identificou que o fator ambiental que mais contribuiu para a variação no SI foi a coabitação, indivíduos que compartilham o mesmo ambiente tendem a ter 50% menos variação que o observado na população geral. O estudo propõe que as condições locais do ambiente são fatores chave na constituição do sistema imune humano, levando a convergência do perfil imune entre pares. Os autores discutem que tal observação pode estar relacionada com o compartilhamento de hábitos e estilo de vida, que tende a ocorrer entre cônjuges. Ademais, uma explicação também plausível envolve uma convergência em termos da microbiota, bem como exposição a patógenos (SONG *et al.*, 2013; KORT *et al.*, 2014; LAX *et al.*, 2014; ROTHSCILD *et al.*, 2018).

1.5. Origens da variação interpopulacional na RI

Ao se pensar na constituição da RI numa perspectiva populacional, observa-se que há uma heterogeneidade na configuração do SI destas, o que inclusive repercute

em diferentes prevalências de doenças imunomediadas (OKADA *et al.*, 2010). A hipótese da higiene proposta por Strachan (1989) postula que o aumento na prevalência de doenças alérgicas em países desenvolvidos está associado a mudanças no estilo de vida que resultam em uma menor carga infecciosa. O mecanismo imune inicialmente proposto se relacionava com o paradigma Th1/Th2, em que o contato com agentes infecciosos na infância aumentaria a resposta Th1, reduzindo a resposta Th2. Já a diminuição nestas infecções causaria, conseqüentemente, um aumento na resposta Th2, envolvida na patogênese de alergias (ROMAGNANI *et al.*, 2004; GARN *et al.*, 2021). Um dos primeiros pontos de fraqueza de tal hipótese foram as infecções helmínticas, tendo em vista que essas têm como mecanismo de resposta uma elevação de citocinas Th2 também. Neste sentido, observa-se uma relação inversa entre presença de infecções helmínticas e predisposição para o desenvolvimento de alergias (ROMAGNANI *et al.*, 2004; SCUDELLARI; 2017).

A descoberta das células regulatórias (Treg) contribuiu para uma mudança mecanicista na hipótese da higiene. Foi percebido que em locais onde há endemismo de parasitoses, os indivíduos tendem a produzir mais células Treg e citocinas associadas (e.g., IL-10), levando ao desenvolvimento de um sistema imune que não é hiper-reativo (VAN DEN BIGGELAAR *et al.*, 2000; GARN *et al.*, 2021). Outro ponto de falha da hipótese, foi explicar o aumento na prevalência de doenças mediadas pela resposta Th1, como as doenças autoimunes, em países desenvolvidos.

Nos dias de hoje, diversas linhas de evidência dão suporte a uma nova visão, a hipótese dos “velhos amigos”, que além de incorporar um novo paradigma, inclui também uma perspectiva evolutiva. Em 2003, Graham Rook e colaboradores propuseram que a exposição para diversos microrganismos “amigáveis” que coevoluíram com os seres humanos - patógenos não infecciosos - é necessária para treinar e regular o sistema imune para reagir apropriadamente. O principal aspecto dessa nova óptica, é que ela considera como principais atores os micróbios não-patogênicos, e não os infecciosos (como pensava-se anteriormente) (ROOK; BRUNET, 2005). Ademais, alguns aspectos da Higiene da Hipótese permanecem na Hipótese dos Velhos Amigos, é provável que de fato mudanças relacionadas a saneamento, alimentação e acesso a água tratada tenham reduzido a exposição a esses microrganismos.

Para além da interação ambiente- indivíduo, uma visão ainda mais abrangente deve incluir também aspectos genéticos, tendo em vista que as populações distintas possuem uma constituição genética também distinta (*i.e.*, diferenças nas frequências alélicas e genóticas) (FREW, 2019). Essa visão mais abrangente é necessária, pois, o *background* genético também contribui para a configuração do seu SI, assim como esse patrimônio genético é resultante da pressão evolutiva exercida ao longo do tempo. Assim, os dados de genética de populações são encarados como uma ferramenta essencial para a melhor compreensão sobre a hipótese da higiene/velhos amigos (SIRONI; CLERICI, 2010).

Uma possibilidade alternativa também postulada, é que muitas variantes genéticas associadas ao *status* imune teriam se comportado de maneira parcialmente neutra, dada a ausência de fatores ambientais que predisõem o desenvolvimento de um SI mais inflamatório, por exemplo. Tais fatores ambientais relacionados a abundância de patógenos, microrganismos não prejudiciais e os “velhos amigos”, que teriam maior efeito na modulação do SI. Na então ausência destes, há a revelação de muitos alelos no genoma humano de que já foram “equilibrados” ou quase neutros, mas são prejudiciais no contexto atual (SIRONI; CLERICI, 2010). Além disso, uma visão ainda mais ampla, denominada hipótese da Biodiversidade, considera o contato com a natureza importante para que haja o desenvolvimento correto do SI (HAAHTELA; 2019). A Figura 3 fornece uma visão geral da hipótese da Biodiversidade/Velhos amigos. Mais informações sobre a complexa interação entre a ancestralidade genética, desequilíbrios imunológicos e infecções (ou a ausência delas) necessitam de grandes esforços para que haja a elucidação das relações entre genótipo/interações microbianas/fenótipo imune em nível molecular.

1.6. Evolução humana e a arquitetura genética da resposta imune

Os agentes patogênicos são considerados uma das principais fontes de pressão seletiva que a humanidade já enfrentou ao longo da sua evolução. Tal fenômeno pode ser observado por meio da investigação dos *loci* genéticos com assinatura de pressão seletiva, em que uma parte considerável destes são variações funcionais em genes relacionados ao SI (FUMAGALLI *et al.*, 2011; KARLSSON *et al.*, 2014; VATSIUO *et al.*, 2016). Neste sentido, pode-se dizer que o *background* genético

relacionado às variações no SI é produto de processos evolutivos que atuaram ao longo da história humana.

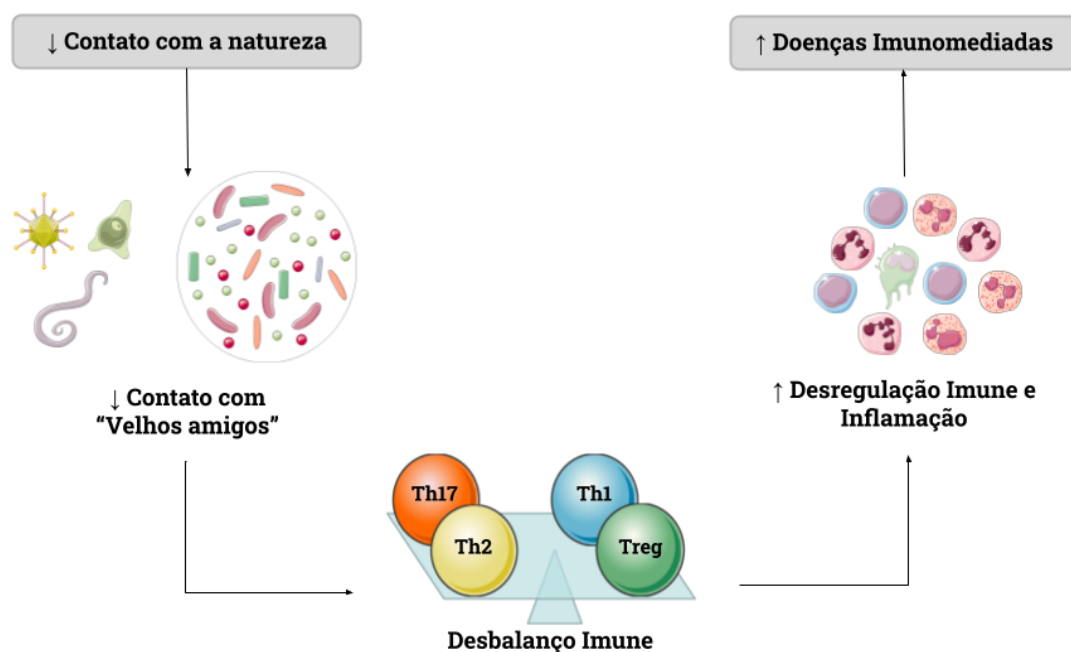


Figura 3. Hipótese da Biodiversidade/Velhos amigos e o Sistema Imune. Mudanças recentes no estilo de vida, maior higiene, menor contato com a biodiversidade, dieta, impactam no contato com os microrganismos e outros organismos importantes para o desenvolvimento correto do SI. O desbalanço imune impacta no desenvolvimento de perfis inflamatórios que estão ligados a doenças imunomediadas.

Considerando o modelo *Out of Africa* (OOA), os humanos anatomicamente modernos surgiram na África há aproximadamente 200 mil anos, e a cerca de 100 mil anos iniciaram-se fluxos migratórios para outras regiões. Os humanos ocuparam a Eurásia, e mais ainda mais recente foi a chegada às Américas, que ocorreu entre 35 e 15 mil anos atrás (NIELSEN *et al.*, 2017; BENTON *et al.*, 2021). Na perspectiva do modelo OOA, é proposto que conforme os humanos modernos migraram para fora da África e ocuparam novos locais, mudanças no ambiente ao qual estavam expostos também ocorreram. Tais mudanças, aconteceram, em especial, no contexto de exposição microbiana/situações patogênicas, resultando em seleção de imunofenótipos específicos (SANZ *et al.*, 2018). Desta forma, a migração dos ancestrais humanos para fora da África implicou na exposição a diferentes tipos de doenças infecciosas, favorecendo então a seleção de variantes genéticas ligadas a uma configuração do

SI mais efetiva àquele contexto patogênico (FUMAGALLI *et al.*, 2011; KARLSSON *et al.*, 2014; VATSIUO *et al.*, 2016). A Figura 4 ilustra o modelo migratório humano, mostrando também a abundância de patógenos e dados epidemiológicos de algumas doenças.

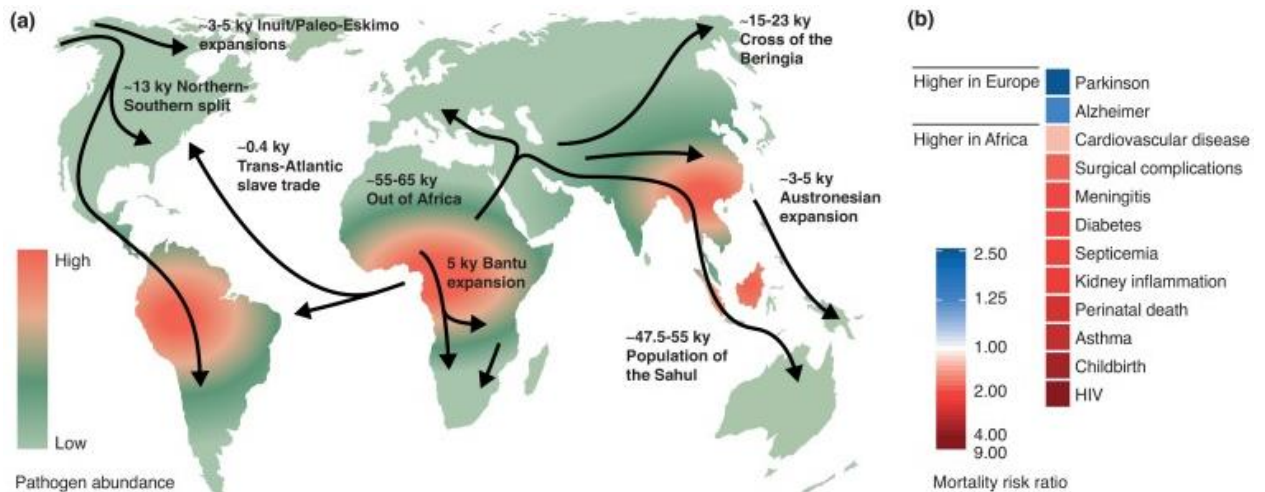


Figura 4. Principais rotas migratórias seguidas pelos ancestrais humanos que se originaram na África. Mapa mostra também a diferença na abundância de patógenos entre regiões. Sanz *et al.*, 2018

O sistema imune humano pode ser considerado, assim, um produto de diversas adaptações resultantes da seleção de variantes funcionais em resposta às interações ambiente-patógeno-hospedeiro ao longo de milhares de anos (DOMÍNGUEZ-ANDRÉS; NETEA, 2019). Apesar de variantes genéticas funcionais potencialmente aumentarem o *fitness* em um dado contexto, também podem impactar diretamente no desenvolvimento de doenças imunomediadas ou traços imunológicos desfavoráveis (CRESPI *et al.*, 2015). Sendo assim, interações entre microrganismos e humanos que ocorreram há milênios são importantes não apenas na resposta contra ameaças patogênicas modernas, mas também no surgimento de doenças autoimunes e inflamatórias observadas nas populações atuais (DOMÍNGUEZ-ANDRÉS; NETEA, 2019).

Um aspecto fundamental da evolução humana é que ela não ocorre de maneira que o fenótipo selecionado é o mais saudável, mas sim o fenótipo que favoreça a sobrevivência do indivíduo, ainda que haja custos. Assim, o conceito de *trade-off* pode ser aplicado, em que se analisa o balanço custo-benefício na seleção de uma dada característica (CRESPI *et al.*, 2015). A principal consequência disso é que, caso a pressão seletiva deixe de existir e não haja tempo o suficiente (*i.e.*, milhares de anos)

para que ocorram mudanças genéticas, existiram variantes nas populações que terão um custo que supera os benefícios nesse novo cenário. Compreender como a seleção natural moldou a variabilidade do genoma representa uma abordagem poderosa para identificar genes que desempenharam um papel importante na sobrevivência do hospedeiro, complementando estudos genéticos imunológicos, clínicos e epidemiológicos (QUACH *et al.*, 2016)

1.7. Ancestralidade genética, miscigenação e produção de citocinas

Estudos recentes sugerem que de fato existe uma relação entre ancestralidade e resposta imune, com indivíduos de ascendência africana demonstrando respostas mais fortes (NÉDÉLEC *et al.*, 2016; QUACH *et al.*, 2016; YAO *et al.*, 2018). Isso pode ser o resultado de processos seletivos em resposta a novos ambientes para as populações europeias, ou uma carga maior de patógenos na África, agora levando a um maior número de distúrbios inflamatórios e autoimunes (BENTON *et al.*, 2021). Por exemplo, Fumagalli e colaboradores identificaram variantes que se correlacionam com a diversidade de helmintos, maior parte delas envolvidas com aspectos da imunidade (2010).

A relação entre o *background* genético e traços imunes se torna ainda mais complexa no contexto de populações miscigenadas, como as latino americanas, pois o genoma destas populações pode ser considerado um mosaico do patrimônio genético das populações ancestrais. Estudos nessa perspectiva podem investigar tanto o impacto da ancestralidade genômica global, assim como o impacto da ancestralidade local, de uma região do genoma do indivíduo, na expressão de imunofenótipos. A figura 5 ilustra a diferença entre ancestralidade genética global e local em indivíduos de uma população.

Nesse sentido, alguns estudos apresentam evidências de diferenças interpopulacionais, ligadas à ancestralidade genética, na resposta imune. Foi demonstrado, por exemplo, que há diferença na expressão gênica em células imunes estimuladas associada com a ancestralidade do indivíduo (Africana ou Europeia) (Nedelec *et al.*, 2016; Quach *et al.*, 2016). Além disso, parte desta diferença interpopulacional pode ser explicada a partir da genética, mais especificamente pela

presença de variantes regulatórias (Cis- e Trans-Acting Expression Quantitative Trait Loci) (Quach *et al.*, 2016). Ademais, tem sido demonstrado um excesso de ancestralidade africana para alguns *loci* em populações recentes da América Latina (Rishishwar *et al.*, 2015; Zhou *et al.*, 2016).

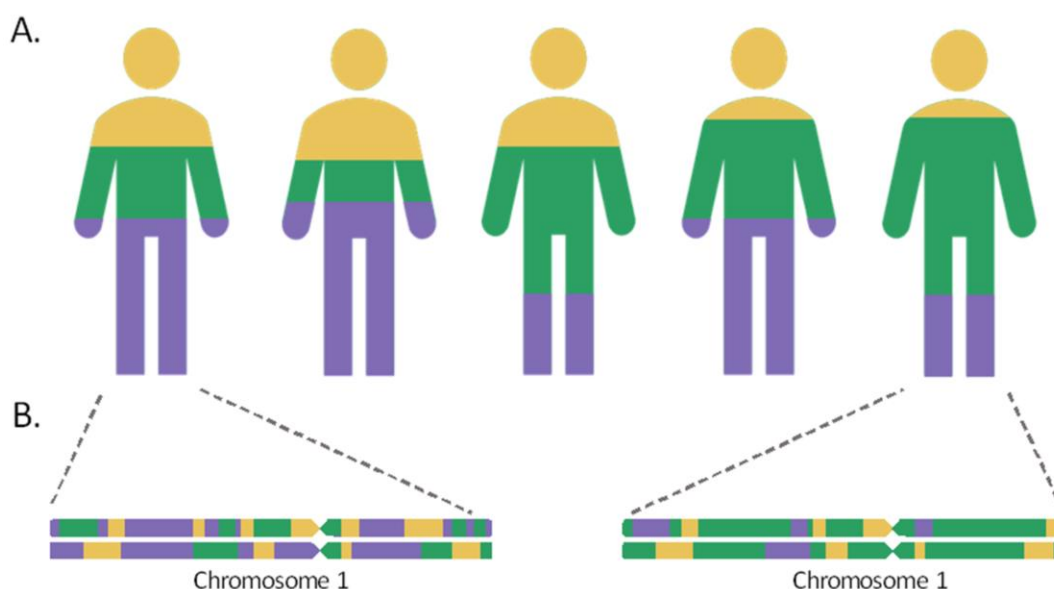


Figura 5. Diferença entre ancestralidade genética global (A) e local (B) em uma população miscigenada. A figura ilustra a composição genética dos indivíduos em uma população recém miscigenada a partir de três populações ancestrais, como a brasileira. SUAREZ-PAJES *et al.*, (2021).

Nessa perspectiva, já foi notado que frequência de alelos em genes ligados à resposta Th2 diferem mais que o esperado entre populações Africanas e Europeias, considerando a diferença genômica, ressaltando que a resposta Th2 está relacionada a defesa contra helmintos (RUSSEL *et al.*, 2015). Nessa mesma linha, ao comparar a expansão *in vitro* de células T, foi observado que células CD4 + IL-4 e IL-13 expandiram-se apenas em adultos africanos (WILFING; 2001). Ademais, o paradigma hoje considera a existência de outros perfis Th, que não necessariamente são contra-regulados. Assim, ao avaliar expressão gênica durante a ativação de células Th em uma coorte de 348 indivíduos saudáveis de três ancestralidades (Africana, Asiática e Europeia), foi notado que a ascendência dos doadores influenciou marcadamente as respostas das células T, com o perfil Th17-*high* sendo associado à ascendência africana (YE *et al.*, 2014). Ademais, mais estudos e mais evidências são necessárias

para que haja um maior suporte para a proposição, para que conclusões válidas possam ser obtidas.

Ainda no contexto da produção de citocinas, Yao *et al.* (2018) investigaram a associação entre ancestralidade genética e a produção de citocinas em que identificaram, por meio da abordagem de mapeamento por miscigenação, a associação entre a variante rs2814778 e os níveis das quimiocinas CCL2 e CCL11. Neste mesmo sentido, Norris *et al.* (2018) investigaram a ancestralidade genética na América Latina e observaram que SNVs ligados à ancestralidade enriquecidos nas populações estudadas estavam relacionados a vias imunológicas, como as de interação citocina-receptor. Além disso, foi observado que alguns destes SNVs relacionados à ancestralidade africana influenciam na expressão do gene *IL10*. Nenhum estudo foi realizado no Brasil utilizando tais estratégias em tais contextos.

1.8. Justificativa

Apesar da importância reconhecida dos estudos imunogenéticos, a maior parte destes ainda se concentram em populações europeias, limitando o potencial de identificar variações genéticas associadas à resposta imune em outras populações (PENG *et al.*, 2021). Neste sentido, o presente trabalho busca investigar a associação entre ancestralidade genética e a produção espontânea das citocinas IFN- γ , IL-5, IL-13 e IL-10 em uma população brasileira de crianças miscigenadas. O estudo tem como alvo as referidas citocinas tendo em vista que estão envolvidas nas respostas do tipo Th1, Th2 e Treg. Nesta perspectiva, foi realizada a análise de associação com ancestralidade global e a abordagem de mapeamento por miscigenação, a fim de identificar variantes genéticas associadas.

2. CAPÍTULO ÚNICO

Admixture mapping approach reveals an association between genomic region 16q23.3 with a lower IL-13 production in a Brazilian pediatric population

Hatilla dos Santos Silva ¹, Helena Mariana Pitangueira Teixeira ², Ana Paula Castro Melo ², Carolina Silva Carvalho ¹, Rafael Pereira Tou ¹, Camila Zolini ¹, Ryan dos Santos Costa ², Mauricio Lima Barreto ³, Eduardo Tarazona-Santos ^{1#}, Camila Alexandrina Viana Figueiredo ^{2#}

¹ Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil.

² Instituto de Ciências da Saúde, Universidade Federal da Bahia (UFBA), Salvador, BA, Brazil

³ Centro de Integração de dados e Conhecimentos para Saúde (CIDACS), Fiocruz, Salvador, BA, Brazil

#These authors contributed equally

Abstract

T-helper cells can be classified according to the cytokine profile produced, characterizing different types of immune response, for example Th1 (IFN- γ), Th2 (Interleukin (IL)-5 and IL-13) and Treg (IL-10) responses. The immune response of an individual is resulted from interaction between environmental, genetic and epigenetic factors. Such genetic components were shaped by exposure to infectious pathogens, and also non-pathogenic microorganisms, faced by humans throughout their evolutionary history. Latin American populations were formed by the admixture process of three ancestral populations that significantly differs in the environment from which they were exposed. Thus, it is important to understand how such admixture processes impact in shaping the immune system of recent admixed populations. Therefore, the present study aims to evaluate the influence of both genomic ancestry on the production of IFN- γ (Th1), IL-5 (Th2), IL-13 (Th2) and IL-10 (Treg) cytokines in a Brazilian population. We observed significant differences for IL-13 production, where individuals with a higher African ancestry were more likely to produce lower levels of IL-13. We further investigated IL-13 production by using the admixture mapping approach. Through admixture mapping, one suggestive signal was pinpointed to local European ancestry at 16q23.3, with fine-mapping analysis refined to one variant in *PLCG2* (rs12598402) and two variants in *CDH13* (rs28376615 and rs75929636). In opposition to the genomic global ancestry pattern, alleles linked to a greater chance of producing IL-13 were more common in individuals with higher African ancestry. Further studies are needed to investigate how immune differences driven by genetic ancestry could predispose to immune mediated diseases.

Keywords: Immunogenetics, Ancestry, Cytokine, IL-13, Th2

1. Introduction

Cytokines are a large group of secreted proteins, with varied structures and functions, involved in the coordination of immune response (IR). T-helper (Th) cells can be differentiated into subpopulations according to their profile of cytokines produced and are involved in different IR contexts, such as Th1 (e.g., IFN- γ), Th2 (e.g., Interleukin (IL)- 5 and IL-13) and Treg (e.g., IL-10) profiles (Eizenberg-Magar, 2017; Martinez-Sanchez *et al.*, 2018). Each profile is linked to distinct scenarios of inflammation, such as allergies and helminth infection (Th2), virus infections and autoimmune diseases (Th1) but also regulatory and tolerance-induced mechanisms (Treg). Although these profiles were the first to be described and studied, other profiles have been recently described, such as Th9, Th17, and Th22, involved in a range of clinical situations. (Luckheeram *et al.*, 2012; Zhu *et al.*, 2019). For example, the Th17 profile is involved in the pathogenesis of both autoimmune diseases and some subtypes of asthma (Bedoya *et al.*, 2013; Zhao *et al.*, 2013).

The immune system (IS) can be characterized by interindividual and interpopulation variability, in which, SI is highly variable among individuals but tends to remain relatively stable over time in a given person (Brodin & Davis, 2017). Such variability has been explained through a multifactorial perspective, which considers genetic, epigenetic, environmental, and evolutionary aspects (Barreiro & Quintana-Murci, 2020). The proportion by which different factors influence cytokine production depends on the marker investigated, especially, its biological role (Brodin *et al.*, 2015). In a study with twins, it was shown, for example, that the estimated heritability for the basal production of IFN- γ , IL-5, IL-13, and IL-10 was 0.46, 0.59, 0.48, and 0.22, respectively (Brodin *et al.*, 2015).

The genetic background related to IS variations in is a product of evolutionary processes that acted throughout human history. As modern humans migrated out of Africa, the environment to which they were exposed changed, in particular, regard to microbial exposure, resulting in the selection of specific immunophenotypes (Sanz *et al.*, 2018). Thus, the environment-pathogen-host relationship contributed to the selection of favorable genetic variants in the given context for human populations (Domínguez-Andrés & Netea, 2019).

Considering the rapid change in the lifestyle of modern humans with a progressive reduction in exposure to pathogens, previously selected variants may confer, in this new context, greater susceptibility to inflammatory diseases. The increase in the incidence and prevalence of autoimmune and allergic diseases has been observed in different populations (Okada *et al.*, 2010) and is explained from this perspective considering the Hygiene Hypothesis (Strachan, 1989), or the more refined Old Friends Hypothesis (Rook *et al.*, 2003). In this sense, it is stated that improvements in human living conditions concerning hygiene, and consequent decrease in exposure to microorganisms, contribute to IR patterns that increase susceptibility to inflammatory diseases. In the same sense, the Biodiversity Hypothesis proposes that the diversity of organisms in the environment contributes positively by promoting a balance in IR that protects the individual from inflammatory diseases (Haahtela, 2019).

Ancestral populations were exposed to different environments; thus, it is assumed that there was a selection of different genetic variants, which could explain the differences in the prevalence of immune-mediated diseases between currently observed populations. More recent evolutionary processes, involving admixture of ancestral populations may also play a role in the constitution of IS genetic architecture (Sanz *et al.*, 2018). For example, an excess of African ancestry has been demonstrated for some *loci* in recent Latin American populations (Rishishwar *et al.*, 2015; Zhou *et al.*, 2016). Furthermore, studies on immune cells have shown that there is a gene expression pattern associated with ancestry in stimulated immune cells (Nedelec *et al.*, 2016; Quach *et al.*, 2016). Furthermore, part of this interpopulation difference could be explained by genetics, more specifically by the presence of regulatory variants (Cis- and Trans-Acting Expression Quantitative Trait Loci) (Quach *et al.*, 2016).

In the context of cytokine production, Yao *et al.* (2018) investigated the association between genetic ancestry and cytokine production, and identified, through the admixture mapping approach, the association between the rs2814778 and the levels of the chemokines *CCL2* and *CCL11*. In addition, Norris *et al.* (2018) investigated genetic ancestry in Latin America and observed that ancestry-linked SNVs enriched in the populations studied were related to immunological pathways, such as cytokine-receptor interaction. Furthermore, it was observed that some of these SNVs related to African ancestry influence the expression of the *IL10* gene. Despite the recognized importance of immunogenetic studies, most of still focus on European

populations, limiting the potential to identify genetic variations associated with the immune response in other populations (Peng *et al.*, 2021). Thus, the present study aims to investigate the association between genetic ancestry and the spontaneous production of IFN- γ , IL-5, IL-13, and IL-10 cytokines in a Brazilian admixed children population. In this perspective, the analysis of association with global ancestry and the admixture mapping approach was carried out to identify associated genetic variants.

2. Methods

Population

Study participants are part of the SCAALA (Social Change Asthma and Allergy in Latin America) program in Salvador – Bahia, Brazil. The detailed study design was previously described in Barreto *et al.* (2006) and Alcântara-Neves *et al.* (2008). In summary, the study population consists of 1,445 children (4 to 11 years old) from the city of Salvador, who were initially recruited to prospectively assess the impact of sanitation on child morbidity. This cohort is suitable for this study since data collection was not biased to recruit a specific phenotype, instead, it is a population-based study representative of Salvador city, later on, used to describe the prevalence of some conditions, such as asthma and allergy.

Information related to socio-demographic and domestic environment variables in 2005 was collected through interviews and questionnaires applied to legal guardians. In addition, whole blood samples were obtained for DNA extraction and cytokine measurement (Barreto *et al.*, 2006; Figueiredo *et al.*, 2009). A consent form with detailed information about the project was obtained from the participant's legal guardian. The project was approved by the ethics committee of the Federal University of Bahia (registration 003-05 / CEP-ISC) and also by the National Council for Ethics in Research (CONEP, resolution nº 15 895/2011).

Cell culture and cytokine production

Detailed information regarding cell culture and subsequent cytokine measurement was previously described by Figueiredo *et al.* (2009). In summary,

venous blood was collected in heparinized tubes and whole blood cells were cultured at a dilution of 1:4 in RPMI medium. Cultured cells were maintained in a humidified environment, 5% CO₂ at 37 °C for 24 hours for the detection of IL-10 and 5 days for the detection of IL -13, IL-5, and IFN- γ , according to the standardization protocol performed.

Spontaneous production of cytokines IFN- γ , IL-5, IL-13, and IL-10 was measured from supernatants using commercial sandwich ELISA kits following the manufacturer's recommended protocol (BD Biosciences Pharmingen, San Diego, CA, USA). Cytokine concentrations were determined using standard curves. The detection limits (lower/upper) for each cytokine were as follows: 18.5/300 pg/mL; 15.63/500 pg/ml; 62.5/4,000 pg/ml; and 31.25/500 pg/ml, for IFN- γ , IL-5, IL-13 and IL-10, respectively. Since a significant proportion of individuals did not produce detectable levels of studied cytokines (65.6% - 94.5%), we categorized them into non-producers and producers, according to the lower detection limits of each assay.

DNA extraction and genotyping

Genomic DNA (gDNA) was extracted from whole blood samples using a commercial kit, following the manufacturer's recommended protocol (Flexigene DNA Kit, Qiagen, Hilden, Germany). SCAALA is one of the three cohorts participating in the EPIGEN-Brasil initiative (Kehdy *et al.*, 2015), which performed the genotyping of 6,487 Brazilian individuals. Among the 1,445 individuals in the SCAALA program, 1309 were genotyped for approximately 2.5 million markers using the Illumina Omni 2.5M chip (San Diego, California). 1,246 unrelated individuals were included in the posterior analysis, due to the removal of possible relationships determined by kinship coefficients (second-degree cutoff; ≥ 0.1).

Genomic global and local ancestry estimation

Global ancestry estimate was performed from a set of 720,438 independent ($r^2 < 0.40$) SNVs (Single Nucleotide Variants) shared between the HapMap project, Human Genome Diversity Project (HDGP), and the populations included in EPIGEN-

Brazil. The proportion of African, European, or Native American ancestry was estimated using the ADMIXTURE software from a tri-hybrid model ($k = 3$). For local ancestry, individual genetic data were phased using the software SHAPEIT2 (Delaneau *et al.* 2011) and subsequently used to perform the local ancestry inference in RFMix software v 1.5.4 (Stanford University, Stanford, CA, USA) (Maples *et al.*, 2013).

The parental population included data from 1000 Genomes Project and HDGP European populations; Africans from four regions in addition to 1000 Genomes Project and HDGP African populations; and Native American populations from Tarazona-Santos group Laboratory of Human Genetic Diversity (LDGH) and Peruvian National Health Institute–INS, Lima, Peru) (Borda *et al.*, 2020) in addition to 1000 Genomes Project and HDGP Native American populations. Detailed information regarding parental populations is provided in Table S1.

RFMix uses a conditional random field parameterized by random forests trained on reference panels, the algorithm can learn from the admixed samples to autocorrect phasing errors and boost local ancestry inferences. To run RFMix, we fixed the number of generations since the admixture event (parameter-G) to 20 (~500 years) and the number of trees to generate per random forest (parameter-t) to 500. Inferences were performed in window lengths (parameter -w) of 0.2 cM. All other parameters present in RFMix were set as default.

Relationship between cytokine production and individual ancestry

The effect of individual ancestry (European, African, and Native American) on cytokine production was assessed by using multivariate binary logistic regression models in the R platform. The analyzes were performed with two models, model 1, adjusting only for age and sex, and model 2, adjusting for other variables that remained significant using the backward selection method. Odds Ratio (OR) and its confidence intervals were estimated considering the 10% increase in the proportion of ancestry considering the formula $e(\beta*10)$, where β represents the coefficient obtained in the logistic regression. Also, we performed the analysis by categorizing individuals into

quartiles of African and European ancestries and median-based for Native American ancestry.

Admixture mapping

We performed admixture mapping using local inferences generated by RFMix. Using logistic regression, we tested the association between IL-13 production status (non-producers vs producers) and each local ancestry (African, European, and Native American). All analyses were adjusted by age, sex, street paving condition, helminth infection, and presence of mold in the house, in addition to the global African ancestry, which would be an indicator of socioeconomic conditions. An additive model was used, which considers the number of ancestry copies (0, 1, or 2) carried by individuals at each window. We used PLINK 1.9 software for regression analysis.

In order to establish a significance threshold accounting for multiple testing, we performed the method proposed by Shriner *et al.* (2011) to estimate the effective number of tests (ENT). The method fits an autoregressive model to each vector of local ancestry for each chromosome for each individual and evaluates the spectral density at frequency zero with the package coda for R. We used the estimates to obtain the Bonferroni-adjusted p-value threshold as 0.05 divided by ENT. We defined a significant p-value $< 2.29 \times 10^{-4}$ (for African ancestry); p-value $< 2.26 \times 10^{-4}$ (for European ancestry) and p-value $< 1.28 \times 10^{-4}$ (for Native American ancestry). We used a p-value $< 1 \times 10^{-3}$ as the suggestive threshold.

Imputation

The imputation was performed through the Michigan Imputation Server (<https://imputationserver.sph.umich.edu>) by incorporating the Consortium on Asthma among African-ancestry (CAAPA - GRCh37/hg19) as the reference panel. The following post-imputation QC (quality control) filters were applied: Non-biallelic SNVs, HWE (10^{-4}), MAF (10^{-4}), and the imputation score R_{sq} of 0.1. The imputed SNVs underwent a second quality control, using as an exclusion parameter the R_{sq} of < 0.80 , a call rate of < 0.95 , and the MAC (minor allele count) < 10 . After the imputation QC,

about 22.6M SNVs remained for further analysis. Imputed and genotyped data were further used to perform the fine-mapping analysis.

Fine-mapping

The significant or suggestive admixture mapping peaks were followed up for fine-mapping analysis focusing on ± 1 Mb centered in the most significant window of each admixture mapping hit. The genotyped and imputed genotypes were tested for association with IL-13 production with the same logistic regression models used in local ancestry analysis. We excluded SNVs with minor allele frequency (MAF) < 0.005 and minor allele count (MAC) ≤ 20 for these analyses. After QC, the regions 3p21.2-3p21.1 (7,726 variants), 10q21.3(11,487 variants) 11q14.1 (8,690 variants), 15q25.1 (10,488 variants), 16q23.3 (15,150 variants), and 17p13.2 (10,674 variants) remained.

We considered the p-values of the respective admixture mapping peaks as the significant threshold for SNV associations. Ultimately, SNVs with significant p-values in the SCAALA cohort were tested using genotyped and imputed data from a Brazilian replication cohort (PROAR Asthma Cohort) following the multiple logistic regression analysis including age, sex, and African genomic ancestry as covariates.

3. Results

Tables S2-5 summarizes the sociodemographic characteristics and other variables according to cytokine production - IFN- γ , IL-5, IL-13, and IL-10 - of the studied population. IFN- γ production was associated with weight at birth and day-care attendance in early childhood. IL-5 production was significantly related to the presence of mold in the house. No significant association was found for IL-13. IL-10 production was associated with the absence of proper paving, tap water, and sewage system conditions. Furthermore, IL-10 was related to infection burden and helminth infection.

In our population, African, European, and Native American median ancestries correspond to 50.29%, 42.83%, and 5.62%, respectively. Results of the association between biogeographical genomic ancestry and cytokine production are shown in Table 1. The cytokines IFN- γ , IL-5 and IL-10 were not associated with genomic

ancestry. IL-13 production was negatively associated with the increasing proportion (each 10%) of African ancestry (OR: 0.93; 95% CI 0.75 - 0.91; $p < 0.001$) in the model adjusted for age, sex and other covariates. Also, when comparing quartiles of ancestry, individuals in the 4th quartile have a greater chance of being a non-producer than individuals in the 1st quartile (OR: 0.49; IC 95% 0.34 - 0.72; $p < 0.001$). European genomic ancestry was also associated with IL-13 production, but in the opposite direction to African ancestry. Moreover, the quantile-based analysis revealed that individuals from second to fourth quartiles had a greater chance to produce the cytokine compared to the first quartile.

Table 1. Association of African ancestry proportion and quartiles with cytokines production. Analysis obtained by logistic regression for the SCAALA. ^aOR for each 10% increase in ancestry.

Variable	African ancestry		European ancestry		Native American ancestry	
	OR (CI 95%)	p-value	OR (CI 95%)	p-value	OR (CI 95%)	p-value
IFN-γ (654 non-producers/ 92 producers)						
Genomic ancestry ^a	1.08 (0.93 - 1.26)	0.32	0.90 (0.77 - 1.07)	0.25	1.12 (0.59 - 2.12)	0.73
Genomic ancestry quantile						
1st	Reference		Reference	-	Reference	-
2nd	1.01 (0.53 - 1.92)	0.98	0.84 (0.41 - 1.27)	0.88	0.98 (0.63 - 1.52)	0.93
3rd	0.94 (0.50 - 1.79)	0.85	0.54 (0.42 - 1.38)	0.16	-	-
4th	1.40 (0.77 - 2.53)	0.27	0.71 (0.30 - 1.39)	0.41	-	-
IL-5 (858 non-producers/ 51 producers)						
Genomic ancestry ^a	0.87 (0.72 - 1.06)	0.16	1.19 (0.97 - 1.45)	0.10	0.73 (0.28 - 1.94)	0.53
Genomic ancestry quantile						
1st	Reference	-	Reference	-	Reference	-
2nd	0.62 (0.28 - 1.37)	0.24	0.77 (0.32 - 1.89)	0.57	1.08 (0.61 - 1.92)	0.79
3rd	0.51 (0.22 - 1.16)	0.11	0.97 (0.42 - 2.28)	0.95	-	-
4th	0.70 (0.33 - 1.49)	0.36	1.62 (0.76 - 3.46)	0.21	-	-
IL-13 (623 non-producers/ 327 producers)						
Genomic ancestry ^a	0.83 (0.75 - 0.91)	< 0.001	1.22 (1.10 - 1.35)	< 0.001	1.31 (0.87 - 1.97)	0.19
Genomic ancestry quantile						
1st	Reference	-	Reference	-	Reference	-
2nd	0.85 (0.59 - 1.24)	0.40	1.50 (1.01 - 2.23)	0.04	1.30 (0.99 - 1.71)	0.06
3rd	0.71 (0.49 - 1.04)	0.08	1.50 (1.01 - 2.23)	0.04		
4th	0.49 (0.34 - 0.72)	< 0.001	2.02 (1.38 - 2.97)	< 0.001		
IL-10 (910 non-producers/ 85 producers)						
Genomic ancestry ^a	0.99 (0.84 - 1.16)	0.88	0.99 (0.83 - 1.18)	0.92	1.40 (0.73 - 2.66)	0.31
Genomic ancestry quantile						
1st	Reference	-	Reference	-	Reference	-
2nd	1.07 (0.55 - 2.07)	0.85	1.49 (0.79 - 2.81)	0.22	1.26 (0.79 - 2.01)	0.33
3rd	1.05 (0.55 - 2.00)	0.88	1.15 (0.59 - 2.25)	0.69	-	-
4th	1.03 (0.53 - 2.01)	0.92	0.90 (0.45 - 1.80)	0.77	-	-

IFN- γ association models were adjusted by age, sex, birth weight, and day care attendance; IL-5 models were adjusted by age, sex, presence of smokers, presence

of mold, paving street and sewage system conditions; IL-13 models were adjusted by age, sex presence of mold, paving street conditions and helminth infections; IL-10 models were adjusted by age, sex, day care attendance, paving street conditions, infection burden and helminth infection.

Since we observed a significant difference in IL-13 production according to the genomic ancestry background, we performed an admixture mapping analysis using IL-13 production as a phenotype in SCAALA population considering the three continental ancestries in Brazilians (African, European, and Native American). Six suggestively associated admixture mapping peaks were found for the IL-13 production in 11q14.1, 16q23.3, 17p13.2 for European ancestry, and 3p21.2 - 3p21.1, 10q21.3, and 15q25.1 for Native American ancestry. Figure 1. shows the Manhattan plots for the admixture mapping results based on RFMix chromosome ancestry inferences and Figures S2-4 show the quantile-quantile plots for such analysis.

No significant or suggestive associations were found for African local ancestry (Table 2). European ancestry in 11q14.1 and 16q23.3 cytogenetic regions were suggestively associated with a lower chance of IL-13 production at detectable levels. In addition, European ancestry in the 17p13.2 region was suggestively associated with a greater chance of IL-13 production (Table 2). Regarding Native American ancestry, 3p21.2 - 3p21.1 regions were related to a lower chance of IL-13 production, while 10q21.3 and 15q25.1 regions were suggestively associated in the opposite direction (Table 2). Table S6. provides statistics of reported suggestively associated peaks for other ancestries regardless of the p-value.

We performed a fine-mapping analysis in order to identify SNVs associated with IL-13 production by analyzing both genotyped and imputed variants in ± 1 Mb centered in the most significant window of each admixture mapping peak. We considered a significant association if the p-value was less than or equal to the correspondent admixture mapping peak.

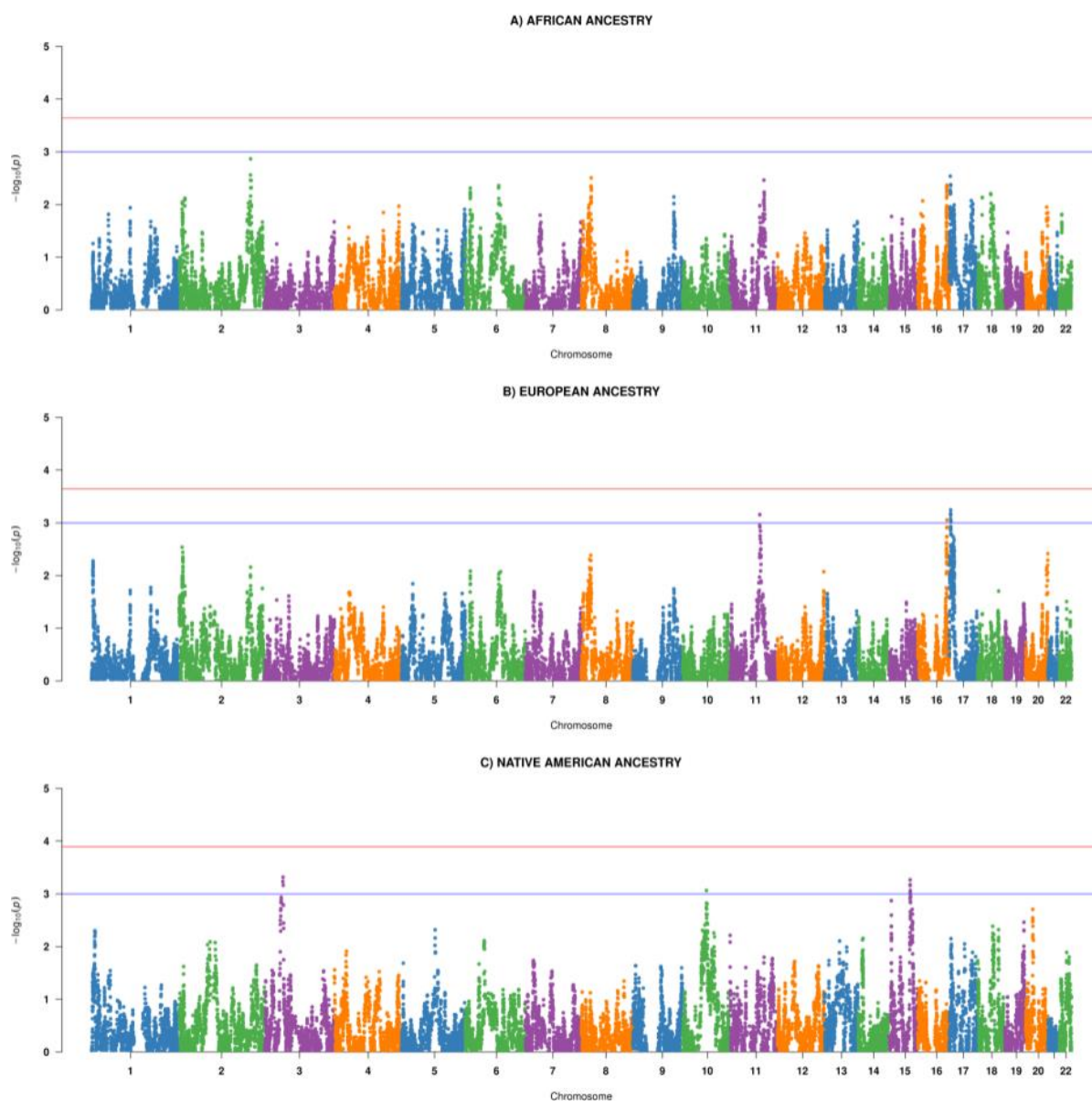


Figure 1. Manhattan plots of the admixture mapping (AM) analyses for IL-13 production. Plot showing the log₁₀ of the p-value (Y-axis) for each local ancestry window (X-axis). No significant association was found for African ancestry AM (A). European ancestry AM analysis shows three suggestive peaks at 11, 16, and 17 chromosomes (B). Native American ancestry AM analysis shows three suggestive peaks at 3, 10, and 15 chromosomes (C).

Table 2. Admixture mapping analysis suggestive peaks for IL-13 production among children from the Social Change, Asthma, Allergy in Latin America (SCAALA) Cohort in Salvador, Brazil.

Ancestry	Chr Region	Chr Position	Initial Window Marker	Final Window Marker	Effect (OR)	P
European	11q14.1	11:84380621 - 84612058	rs17147488	rs7943571	0.72 (0.60 - 0.87)	6.98E-04
	16q23.3	16:82828365 - 82865785	rs12929663	rs9936998	0.73 (0.60 - 0.88)	8.86E-04
		17:3564235 - 3672941	rs3826493	rs11649949	1.39 (1.15 - 1.67)	6.65E-04
		17:3674153 - 3703782	rs12604024	rs4300684	1.38 (1.15 - 1.67)	7.22E-04
	17p13.2	17:3705526 - 3762271	rs2891	rs57724104	1.38 (1.14 - 1.67)	8.53E-04
		17:3763574 - 3794352	rs1558294	rs7216488	1.39 (1.15 - 1.68)	5.70E-04
		17:4455500 - 4490468	rs7206974	rs7210765	1.37 (1.14 - 1.64)	9.54E-04
N. American		3:51368795 - 51962081	rs78870156	rs323881	0.45 (0.29 - 0.71)	5.84E-04
	3p21.2 - 3p21.1	3:51963904 - 52536308	rs11921171	rs758801	0.45 (0.29 - 0.71)	4.84E-04
		3:52544155 - 53178658	rs114506140	rs62254273	0.47 (0.3 - 0.72)	6.88E-04
	10q21.3	10:68276614 - 68396283	rs76422529	rs7072222	1.76 (1.26 - 2.46)	8.61E-04
		15:80563203 - 80637884	rs11856226	rs9920521	1.86 (1.3 - 2.67)	6.70E-04
		15:80639564 - 80754984	rs3848198	rs8028295	1.89 (1.32 - 2.71)	5.42E-04
	15q25.1	15:80756302 - 80878680	rs115471347	rs73494771	1.86 (1.3 - 2.67)	6.94E-04
		15:80879709 - 81077104	rs76810611	rs12372942	1.82 (1.28 - 2.59)	9.45E-04
	15:81077470 - 81104785	rs10519292	rs28468684	1.83 (1.28 - 2.61)	8.59E-04	

We attempted to replicate the findings related to IL-13 production in 653 individuals from the PROAR cohort. Three SNVs, within the 16q23.3 genomic region, previously associated in Fine-mapping to the SCAALA cohort were also associated in the PROAR cohort tested by logistic regression considering a significant p-value < 0.05. The *PLCG2* variant rs12598402-C was associated in both cohorts with a lower level of IL-13 with a similar effect size. The *CDH3* genetic variants rs28376615-C and rs75929636-G were related to a higher production of IL-13 in both the discovery and replication cohort (Table 3).

Table 3. Fine-mapping significant associations for IL-13 production identified through imputed variants. Analysis obtained by logistic regression shows only the associated common SNVs for the SCAALA and ProAr.

Cohort	Region	SNV	Position	A1	A2	MAF	OR (CI 95%)	P	Gene
SCAALA	16q23.3	rs12598402	16:81968848	C	T	0.22	0.67 (0.54 - 0.84)	5.68E-04	<i>PLCG2</i>
		rs28376615	16:83480673	C	A	0.08	1.72 (1.25 - 2.36)	8.54E-04	<i>CDH13</i>
		rs75929636	16:83477556	G	A	0.11	1.62 (1.22 - 2.15)	8.77E-04	<i>CDH13</i>
PROAR	16q23.3	rs12598402	16:81968848	C	T	0.21	0.68 (0.49 - 0.95)	2.43E-02	<i>PLCG2</i>
		rs28376615	16:83480673	C	A	0.09	1.63 (1.09 - 2.44)	1.65E-02	<i>CDH13</i>
		rs75929636	16:83477556	G	A	0.10	1.62 (1.09 - 2.42)	1.83E-02	<i>CDH13</i>

We investigated the relationship between genotype probabilities according to African genomic ancestry (Figure 2). We found a negative correlation between African ancestry and rs12598402-C frequency, which was associated with a lower chance to produce IL-13. Also, rs28376615-C and rs75929636-G frequencies were positively related to African ancestry, both variants were associated with a greater chance of producing detectable levels of IL-13. These results were consistent for European ancestry as well Figure S4, but in the opposite direction.

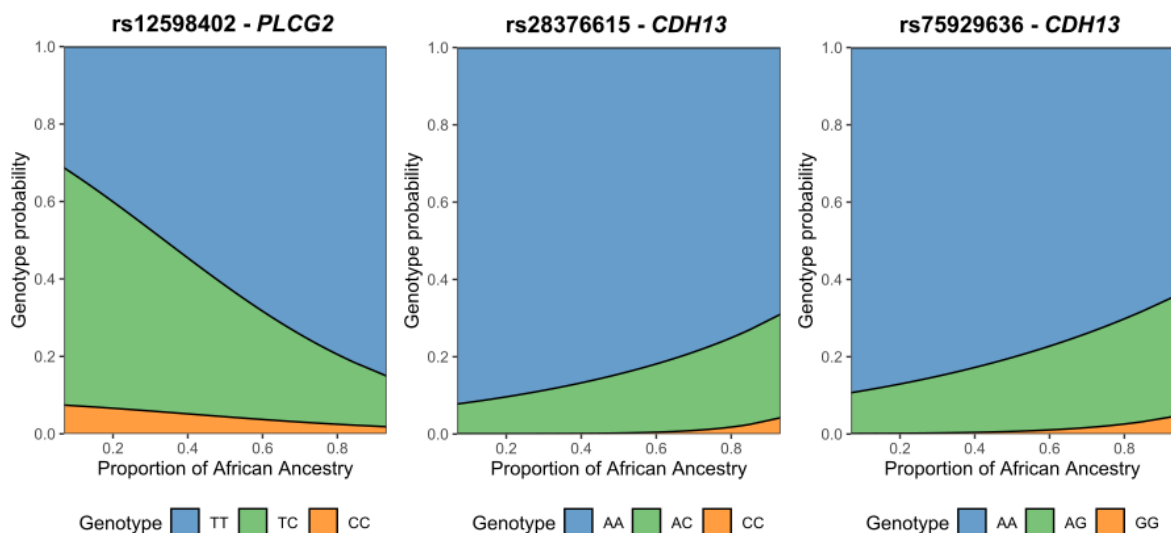


Figure 2. Frequency distribution of genotypes according to the estimated African genomic ancestry for the SCAALA population. Fitted multinomial logistic regression models showing the association between the distribution of genotypes for (a) rs12598402, (b) 28376615, and (c) 75929636 according to African ancestry.

4. Discussion

The primary aim of this study is to describe, according to biogeographical genomic ancestry, the spontaneous cytokine profile of children aged 4 to 11 years in a low-income population from a developing country. In order to correctly adjust the association models, we performed an exploratory analysis considering socio-environmental and demographic variables. Although, Figueiredo *et al.* (2009) previously investigated the effects of environmental exposures in addition to biological characteristics in immune homeostasis for this population.

In the present study, we report a negative association between individual African ancestry and IL-13 production status in childhood. In accordance, we also report positive associations between European ancestry and IL-13 production. To the

best of our knowledge, is one of the first studies to investigate the association between genomic ancestry and IL-13 production in a Latin American population. Interleukin-13 is one of the signature cytokines of type 2 immune inflammation, mainly related to allergen and parasite responses, which can be produced by both innate and adaptive cells (Juntilla, 2018). The IL-13 regulates cellular functions through cell receptors, which downstream signaling involves activating the transcriptional machinery.

The association between individual genomic ancestry and Th2 marker has been little investigated, although there are few related results. For example, in a systematic review, it was shown a negative correlation between Th2-mediated conditions, such as allergy, atopic dermatitis, and allergic rhinoconjunctivitis with socioeconomic level (Uphoff *et al.*, 2015). It should be pointed out that for our population the socioeconomic level (mother education and household income) is associated with African Biogeographic Ancestry (Silva *et al.*, 2018).

Moreover, Yao *et al.*, reported that the mean concentration of IL-4 was 13% higher in European ancestry than in African ancestry women, although it lacks significance when adjusted for educational level (2018). Although they did not investigate IL-13 levels, there is a crosslink between both cytokines, IL-4 and IL-13, in which biological properties are similar. Moreover, in a study conducted in a South African adult population, ethnic groups were compared according to several cytokines production, in addition to other findings, significantly lower levels of IL-13 were found in the black participants compared to whites (Crouch *et al.*, 2020). However, inverse associations are also reported, in which individuals with African ancestry are more likely to develop Th2-mediated conditions (Vergara, 2013; Daya *et al.*, 2019), the association between ancestry and Th2-mediated conditions is not totally conclusive, especially when considering the crucial role of environmental modulator factors.

In our study, the global African ancestry was related to a lower chance of producing IL-13, which is in the opposite direction expected by the hypothesized increased genetic background that predisposes to a Type 2 inflammatory profile in individuals with African ancestry. Although this hypothesis is mainly applicable to a helminth-free environment, therefore it should be further investigated the explanatory power of such affirmation in a country with current helminthiasis (such as Brazil). As previously reported, individuals with higher African biogeographical ancestry are more

likely to live in poor sanitary conditions and more exposed to infectious agents, such as helminths (Silva *et al.*, 2018).

It is possible that longstanding exposure to these features is connected to the development of mechanisms that down-regulate Th2 immunity due to helminth chronic infection. Furthermore, mechanisms that explain such downregulation can be associated with both host and parasite-related factors (Nutman, 2015). Mechanistically, chronic human responses to helminths are related to the expansion of Tregs cells, which involves the production of immunomodulatory cytokines IL-10 and TGF- β (Finlay *et al.*, 2014; Maizels *et al.*, 2018; Maizels, 2020). As previously demonstrated in a study from the same cohort, children living in circumstances of poor hygiene during the first three years of life have elevated IL-10 spontaneous production later in life (Figueiredo *et al.*, 2009; Figueiredo *et al.*, 2010).

In admixture mapping, six regions were associated with IL-13 levels. The subsequent fine-mapping approach revealed for the 16q23.3 region three variants associated with IL-13 production, which was further replicated in an independent cohort. None of the SNVs found here were previously associated with IL-13 production or other makers of type 2 immune response. The rs12598402_C, in the *PLCG2* gene, was associated with a lower chance of producing IL-13. This gene encodes the Phospholipase C gamma 2 (PLC γ 2), an enzyme involved in hydrolyzing phosphatidylinositol 4,5-bisphosphate, producing second messenger molecules diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3) (Neves *et al.*, 2018). The IP3 molecule induces calcium release from the endoplasmic reticulum to the cytoplasm of the cell, and DAG is linked to the activation of cell signaling protein. PLC γ 2 is mainly found in hematopoietic cells, immune cell receptors BCR and Fc receptors, for example, recruit tyrosine kinases and subsequently activate the phospholipase. PLC γ 2 is necessary for the appropriate development of B cells in producing antigen-specific responses (Wang *et al.*, 2000; Jackson *et al.*, 2021). Also, PLC γ 2 is likely implicated in allergic responses, specifically with regard to the mast degranulation mediated by the interaction between IgE and Fc γ R receptor (Wen *et al.*, 2002; Ombrello, 2012).

Noteworthy, it was reported that a PLC inhibitor, and also inhibitors of other downstream molecules, blocked the IL-13 expression and production, indicating that IL-13 expression and production are related to the PLC γ pathway (Park *et al.*, 2009).

Indeed, there are few immune-mediated conditions caused by mutations in the gene, such as cold urticaria, immunodeficiency, and autoimmunity (Martin-Nalda *et al.*, 2020). In a study conducted on southwestern Europeans with North African ancestry influences, whole-exome sequencing analysis revealed an enrichment of likely deleterious variants among asthma cases in the same region, particularly in the phospholipase C γ 2 (*PLCG2*) gene (Guillen-Guio *et al.*, 2020). The variant associated here is identified as an sQTL in whole blood (GTEx) and also as an eQTL in blood cells (eQTLgen), leading to a lower expression of functional *PLCG2*. Further functional studies are needed to confirm the specific mechanisms linking *PLCG2* and its genetic variants with a Th2 response pattern.

The rs28376615_C and rs75929636_G, in the *CDH13* gene, were both associated with a higher chance of producing IL-13. *CDH13* encodes T-cadherin, a non-canonical member of the cadherin superfamily, which lacks the transmembrane and cytoplasmic domains. In a study that compared asthmatic individuals grouped into Th2 low or high response, *CDH13* expression was positively correlated to a Th2 pattern, which translated into correlation with IL13 expression and also clinical features mediated by a Th2 response (Choy *et al.*, 2011). Also, it was observed that knockout mice for *Cdh13* have reduced allergy compared to the Wild Type, in accordance IL-13 levels were also diminished (Williams *et al.*, 2012). In addition to other functions, such as cell adhesion, T-cadherin is one of the receptors for the adipocytokine adiponectin, which indicates that T-cadherin might modulate its activity (Rubina *et al.*, 2021). Indeed, it was shown that T-cad $^{-/-}$ mice have markedly elevated blood adiponectin, which is likely to explain the effect of the absence of T-cadherin in modulating allergy response (Williams *et al.*, 2012).

Adiponectin is a multimeric protein that is present in different biologically active isoforms, largely secreted from adipocytes. Adiponectin serum levels are negatively correlated with the presence of an allergy, in which mechanisms of protection are probably related to the expression of immunomodulatory cytokines, such as IL-10 by Tregs (Hsueh *et al.*, 2010; Ramos-Ramírez *et al.*, 2021). Additionally, SNVs in the *CDH13* gene are consistently associated with adiponectin levels in genome-wide association studies performed in different populations (Chung *et al.*, 2011; Morisaki *et al.*, 2012; Spracklen *et al.*, 2020). We hypothesize that *CDH13* variants that could

impact *CDH13* expression or function, could lead to changes in the adiponectin pathway resulting in differential IL-13 production.

A remarkable finding from our study is that variants associated with a greater chance of IL-13 production are indeed more frequent in African ancestry. Thus, this study reinforces the postulated hypothesis in the African genomic background predisposes to an enhanced Th2 response by the enrichment of variants linked to greater levels of Th2 mediators (Russel *et al.*, 2015). Russel and colleagues (2015) compared Fst values, a population differentiation metric, and found that variants in Th2-related genes differ more between African and European populations compared to the genomic background. Although, we recognize that the expression of such genomic background is influenced by the environmental exposures that can modulate the expression of immune-relevant cytokines.

In conclusion, we found an association between biogeographical genomic ancestry and IL-13 levels, which led us to perform three admixture mapping/fine-mapping for IL-13 production. We provide a list of three candidate ancestry-enriched SNVs that are associated with IL-13 production. Altogether, our results show that indeed variants linked to a greater chance of IL-13 production, a Th2 signature cytokine, are enriched in African ancestry populations and can impact IL-13 production in an admixed population as well.

REFERENCES

1. ALCANTARA-NEVES, Neuza Maria, *et al.* The effect of single and multiple infections on atopy and wheezing in children. *Journal of allergy and clinical immunology*, 2012, 129.2: 359-367. e3.
2. BARRETO, Mauricio L., *et al.* Risk factors and immunological pathways for asthma and other allergic diseases in children: background and methodology of a longitudinal study in a large urban center in Northeastern Brazil (Salvador-SCAALA study). *BMC pulmonary medicine*, 2006, 6.1: 1-10.
3. BEDOYA, Simone Kennedy, *et al.* Th17 cells in immunity and autoimmunity. *Clinical and Developmental Immunology*, 2013, 2013.
4. BRODIN, Petter, *et al.* Variation in the human immune system is largely driven by non-heritable influences. *Cell*, 2015, 160.1-2: 37-47.
5. BRODIN, Petter; DAVIS, Mark M. Human immune system variation. *Nature reviews immunology*, 2017, 17.1: 21-29.

6. BORDA, Víctor *et al.* The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. *Proceedings of the National Academy of Sciences*, v. 117, n. 51, p. 32557-32565, 2020.
7. CAMPOS, Ana Clara P., *et al.* Understanding the dynamics of asthma symptoms between childhood and adolescence using latent transition analysis. *International Journal of Public Health*, 2020, 65.6: 957-967.
8. CHOY, David F. *et al.* Gene expression patterns of Th2 inflammation and intercellular communication in asthmatic airways. *The Journal of Immunology*, v. 186, n. 3, p. 1861-1869, 2011.
9. CHUNG, Chia-Min, *et al.* A genome-wide association study reveals a quantitative trait locus of adiponectin on CDH13 that predicts cardiometabolic outcomes. *Diabetes*, 2011, 60.9: 2417-2423.
10. CROUCH, Simone H. *et al.* Distinct inflammatory mediator patterns in young black and white adults: the African-predict study. *Cytokine*, v. 126, p. 154894, 2020.
11. DAYA, Michelle; BARNES, Kathleen C. African American ancestry contribution to asthma and atopic dermatitis. *Annals of Allergy, Asthma & Immunology*, v. 122, n. 5, p. 456-462, 2019.
12. DELANEAU, Olivier; MARCHINI, Jonathan; ZAGURY, Jean-François. A linear complexity phasing method for thousands of genomes. *Nature methods*, v. 9, n. 2, p. 179-181, 2012.
13. DOMÍNGUEZ-ANDRÉS, Jorge; NETEA, Mihai G. Impact of historic migrations and evolutionary processes on human immunity. *Trends in immunology*, 2019, 40.12: 1105-1119.
14. EIZENBERG-MAGAR, Inbal, *et al.* Diverse continuum of CD4+ T-cell states is determined by hierarchical additive integration of cytokine signals. *Proceedings of the National Academy of Sciences*, 2017, 114.31: E6447-E6456.
15. FIGUEIREDO, Camila Alexandrina, *et al.* Environmental conditions, immunologic phenotypes, atopy, and asthma: new evidence of how the hygiene hypothesis operates in Latin America. *Journal of allergy and clinical immunology*, 2013, 131.4: 1064-1068. e1.
16. FIGUEIREDO, Camila Alexandrina, *et al.* Spontaneous cytokine production in children according to biological characteristics and environmental exposures. *Environmental health perspectives*, 2009, 117.5: 845-849.
17. FIGUEIREDO, Camila Alexandrina *et al.* Chronic intestinal helminth infections are associated with immune hyporesponsiveness and induction of a regulatory network. *Infection and immunity*, v. 78, n. 7, p. 3160-3167, 2010.
18. FINLAY, Conor M.; WALSH, Kevin P.; MILLS, Kingston HG. Induction of regulatory cells by helminth parasites: exploitation for the treatment of inflammatory diseases. *Immunological reviews*, v. 259, n. 1, p. 206-230, 2014.

19. GUILLEN-GUIO, Beatriz *et al.* Admixture mapping of asthma in southwestern Europeans with North African ancestry influences. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, v. 318, n. 5, p. L965-L975, 2020.
20. HAAHTELA, Tari. A biodiversity hypothesis. *Allergy*, 2019, 74.8: 1445-1456.
21. HSUEH, Kai-Chung *et al.* Serum leptin and adiponectin levels correlate with severity of allergic rhinitis. *Pediatric Allergy and Immunology*, v. 21, n. 1-Part-II, p. e155-e159, 2010.
22. JACKSON, Jacob T. *et al.* The role of PLC γ 2 in immunological disorders, cancer, and neurodegeneration. *Journal of Biological Chemistry*, v. 297, n. 2, 2021.
23. JUNTTILA, Ilkka S. Tuning the cytokine responses: an update on interleukin (IL)-4 and IL-13 receptor complexes. *Frontiers in immunology*, v. 9, p. 888, 2018.
24. KEHDY, Fernanda SG, *et al.* Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proceedings of the National Academy of Sciences*, 2015, 112.28: 8696-8701.
25. LEE, Jongdae, *et al.* The Multifaceted Role of Th1, Th9, and Th17 Cells in Immune Checkpoint Inhibition Therapy. *Frontiers in Immunology*, 2021, 12.
26. LIMA-COSTA, Maria Fernanda; FIRMO, Josélia OA; UCHÔA, Elizabeth. The Bambuí Cohort Study of Aging: methodology and health profile of participants at baseline. *Cadernos de saude publica*, 2011, 27: s327-s335.
27. LUCKHEERAM, Rishi Vishal, *et al.* CD4+ T cells: differentiation and functions. *Clinical and developmental immunology*, 2012, 2012.
28. MAIZELS, Rick M.; SMITS, Hermelijn H.; MCSORLEY, Henry J. Modulation of host immunity by helminths: the expanding repertoire of parasite effector molecules. *Immunity*, v. 49, n. 5, p. 801-818, 2018.
29. MAIZELS, Rick M. Regulation of immunity and allergy by helminth parasites. *Allergy*, v. 75, n. 3, p. 524-534, 2020.
30. MAPLES, Brian K. *et al.* RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *The American Journal of Human Genetics*, v. 93, n. 2, p. 278-288, 2013.
31. MARTINEZ-SANCHEZ, Mariana E., *et al.* Role of cytokine combinations on CD4+ T cell differentiation, partial polarization, and plasticity: continuous network modeling approach. *Frontiers in physiology*, 2018, 9: 877.
32. MARTÍN-NALDA, Andrea *et al.* Severe autoinflammatory manifestations and antibody deficiency due to novel hypermorphic PLCG2 mutations. *Journal of clinical immunology*, v. 40, n. 7, p. 987-1000, 2020.

33. MATHIAS, Rasika Ann, *et al.* A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nature communications*, 2016, 7.1: 1-10.
34. MORISAKI, Hiroko, *et al.* CDH13 gene coding t-cadherin influences variations in plasma adiponectin levels in the Japanese population. *Human mutation*, 2012, 33.2: 402-410.
35. NÉDÉLEC, Yohann, *et al.* Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell*, 2016, 167.3: 657-669. e21.
36. NEVES, João Farela *et al.* Novel PLCG2 mutation in a patient with APLAID and cutis laxa. *Frontiers in immunology*, p. 2863, 2018.
37. NORRIS, Emily T., *et al.* Genetic ancestry, admixture and health determinants in Latin America. *BMC genomics*, 2018, 19.8: 75-87.
38. NUTMAN, Thomas B. Looking beyond the induction of Th2 responses to explain immunomodulation by helminths. *Parasite immunology*, v. 37, n. 6, p. 304-313, 2015.
39. OKADA, Hea, *et al.* The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. *Clinical & Experimental Immunology*, 2010, 160.1: 1-9.
40. OMBRELLO, Michael J. *et al.* Cold urticaria, immunodeficiency, and autoimmunity related to PLCG2 deletions. *New England Journal of Medicine*, v. 366, n. 4, p. 330-338, 2012.
41. PARK, Shin-Young, *et al.* House dust mite allergen Der f 2-induced phospholipase D1 activation is critical for the production of interleukin-13 through activating transcription factor-2 activation in human bronchial epithelial cells. *Journal of Biological Chemistry*, 2009, 284.30: 20099-20110.
42. PENG, Kerui, *et al.* Diversity in immunogenomics: the value and the challenge. *Nature Methods*, 2021, 18.6: 588-591.
43. QUACH, Hélène, *et al.* Genetic adaptation and Neandertal admixture shaped the immune system of human populations. *Cell*, 2016, 167.3: 643-656. e17.
44. QUINTANA-MURCI, Lluís. Human immunology through the lens of evolutionary genetics. *Cell*, 2019, 177.1: 184-199.
45. RAMOS-RAMÍREZ, Patricia *et al.* Adiponectin/AdipoR1 Axis Promotes IL-10 Release by Human Regulatory T Cells. *Frontiers in Immunology*, v. 12, 2021.
46. RISHISHWAR, Lavanya, *et al.* Ancestry, admixture and fitness in Colombian genomes. *Scientific reports*, 2015, 5.1: 1-16.
47. ROOK, Graham AW; MARTINELLI, Roberta; BRUNET, Laura Rosa. Innate immune responses to mycobacteria and the downregulation of atopic responses. *Current opinion in allergy and clinical immunology*, 2003, 3.5: 337-342.

48. RUBINA, K. A. *et al.* Revisiting the multiple roles of T-cadherin in health and disease. *European Journal of Cell Biology*, v. 100, n. 7-8, p. 151183, 2021.
49. RUSSELL, Shirley B. *et al.* Pleiotropic effects of immune responses explain variation in the prevalence of fibroproliferative diseases. *PLoS genetics*, v. 11, n. 11, p. e1005568, 2015.
50. SANZ, Joaquin; RANDOLPH, Haley E.; BARREIRO, Luis B. Genetic and evolutionary determinants of human population variation in immune responses. *Current opinion in genetics & development*, 2018, 53: 28-35.
51. SILVA, Thiago Magalhães da *et al.* Biogeographical ancestry is associated with socioenvironmental conditions and infections in a Latin American urban population. *SSM-population health*, v. 4, p. 301-306, 2018.
52. SPRACKLEN, Cassandra N., *et al.* Adiponectin GWAS loci harboring extensive allelic heterogeneity exhibit distinct molecular consequences. *PLoS genetics*, 2020, 16.9: e1009019.
53. STRACHAN, David P. Hay fever, hygiene, and household size. *BMJ: British Medical Journal*, 1989, 299.6710: 1259.
54. UPHOFF, Eleonora *et al.* A systematic review of socioeconomic position in relation to asthma and allergic diseases. *European Respiratory Journal*, v. 46, n. 2, p. 364-374, 2015.
55. VERGARA, Candelaria, *et al.* African ancestry is a risk factor for asthma and high total IgE levels in African admixed populations. *Genetic epidemiology*, 2013, 37.4: 393-401.
56. WANG, Demin, *et al.* Phospholipase C γ 2 is essential in the functions of B cell and several Fc receptors. *Immunity*, 2000, 13.1: 25-35.
57. WEN, Renren, *et al.* Phospholipase C γ 2 is essential for specific functions of Fc ϵ R and Fc γ R. *The Journal of Immunology*, 2002, 169.12: 6743-6752.
58. WILLIAMS, Alison S., *et al.* Role of the adiponectin binding protein, T-cadherin (Cdh13), in allergic airways responses in mice. *PLoS One*, 2012, 7.7: e41088.
59. YAO, Song, *et al.* Genetic ancestry and population differences in levels of inflammatory cytokines in women: role for evolutionary selection and environmental factors. *PLoS genetics*, 2018, 14.6: e1007368.
60. ZHAO, Jingyue; LLOYD, Clare M.; NOBLE, Alistair. Th17 responses in chronic allergic airway inflammation abrogate regulatory T-cell-mediated tolerance and contribute to airway remodeling. *Mucosal immunology*, 2013, 6.2: 335-346.
61. ZHOU, Quan; ZHAO, Liang; GUAN, Yongtao. Strong selection at MHC in Mexicans since admixture. *PLoS genetics*, 2016, 12.2: e1005847.
62. ZHU, Xiaoliang; ZHU, Jinfang. CD4 T helper cell subsets and related human immunological disorders. *International Journal of Molecular Sciences*, 2020, 21.21: 8011.

Supplementary Material

Table S1. Populations used for ancestry inference according to continental origin and data source.

Continental Origin	Population	Source	Individuals (n)
Africans	ACB	1000 Genomes	96
	ASW	1000 Genomes	62
	ESN	1000 Genomes	99
	GWD	1000 Genomes	113
	LWK	1000 Genomes	99
	MSL	1000 Genomes	85
	YRI	1000 Genomes	108
	Biaka	HGDP	20
	Kenya	HGDP	9
	Mandenka	HGDP	18
	Mbuti	HGDP	8
	Namibia	HGDP	2
	South Africa	HGDP	4
	Yoruba	HGDP	18
	Botswana	Crawford <i>et al.</i> 2017	140
	Gahna	Ghana Prostate Study (NCI-NIH) - (Chokkalingam <i>et al.</i> 2012)	968
	Tanzania	Crawford <i>et al.</i> 2017	43
Uganda	EMBLEM	795	
TOTAL			2687
Europeans	CEU	1000 Genomes	99
	GBR	1000 Genomes	91
	IBS	1000 Genomes	107
	TSI	1001 Genomes	107
	Basque	HGDP	21
	Bergamo	HGDP	10
	French	HGDP	24
	Orcadian	HGDP	13
	Sardinian	HGDP	24
	Tuscan	HGDP	6
TOTAL			502
Native Americans	Karitiana	HGDP	8
	Maya	HGDP	19
	Pima	HGDP	11
	Surui	HGDP	6
	Peruvian Natives	LDGH/INS (Borda <i>et al.</i> 2020)	358
TOTAL			402

Table S2. IFN- γ production in SCAALA cohort according to sociodemographic and environmental variables.

Variable		Non-producers	Producers	p-value
Sex	Female	47.66% (376)	41.51% (44)	0.277
	Male	52.34% (413)	58.49% (62)	
Age		6 (5-8)	7 (5-7)	0.398
BMI Z-score		-0.33 (-1.00 - 0.50)	-0.18 (-0.82 - 0.72)	0.077
Birth weight	\geq 2500g	90.51% (706)	97.17% (103)	0.036
	< 2500g	9.49% (74)	2.83% (3)	
Infection Burden	Light	41.54% (324)	48.11% (51)	0.238
	Heavy	58.46% (456)	51.89% (55)	
Mother education	Only primary school	20.9% (163)	20.75% (22)	0.141
	More than primary school	46.67% (364)	55.66% (59)	
	More than high school	32.44% (253)	23.58% (25)	
Income	< 1 MW	52.31% (408)	53.77% (57)	0.961
	1 - 2 MW	31.15% (243)	30.19% (32)	
	>2 MW	16.54% (129)	16.04% (17)	
Day care attendance	Yes	14.1% (110)	22.64% (24)	0.031
	No	85.9% (670)	77.36% (82)	
Smoker in home	Yes	25.13% (196)	24.53% (26)	0.989
	No	74.87% (584)	75.47% (80)	
Mold in home	Yes	69.74% (544)	63.21% (67)	0.21
	No	30.26% (236)	36.79% (39)	
Tap water	Always present	80.9% (631)	76.42% (81)	0.337
	Absent in any time	19.1% (149)	23.58% (25)	
Paving street	Always present	26.15% (204)	28.3% (30)	0.724
	Absent in any time	73.85% (576)	71.7% (76)	
Sewage	Always present	42.56% (332)	41.51% (44)	0.919
	Absent in any time	57.44% (448)	58.49% (62)	
Helminthic infection	Yes	27.05% (211)	29.25% (31)	0.719
	No	72.95% (569)	70.75% (75)	

Table S3. IL-5 production in SCAALA cohort according to sociodemographic and environmental variables.

	Variable	Non-producers	Producers	p-value
Sex	Female	45.72% (481)	51.67% (31)	0.444
	Male	54.28% (571)	48.33% (29)	
	Age	6 (5-8)	6.5 (5-8)	0.252
	BMI Z-score	-0.28 (-0.95 - 0.51)	-0.90 (-0.05 - 0.70)	0.633
Birth weight	≥ 2500g	91.74% (955)	93.33% (56)	0.845
	< 2500g	8.26% (86)	6.67% (4)	
Infection Burden	Light	59.08% (615)	58.33% (35)	1
	Heavy	40.92% (426)	41.67% (25)	
Mother education	Only primary school	22.29% (232)	20% (12)	0.910
	More than primary school	47.65% (496)	48.33% (29)	
	More than high school	30.07% (313)	31.67% (19)	
Income	< 1 MW	54.47% (567)	45% (27)	0.359
	1 - 2 MW	29.2% (304)	35% (21)	
	>2 MW	16.33% (170)	20% (12)	
Day care attendance	Yes	15.66% (163)	15% (9)	1
	No	84.34% (878)	85% (51)	
Smoker in home	Yes	27.38% (285)	20% (12)	0.27
	No	72.62% (756)	80% (48)	
Mold in home	Yes	69.74% (726)	53.33% (32)	0.012
	No	30.26% (315)	46.67% (28)	
Tap water	Always present	79.92% (832)	85% (51)	0.428
	Absent in any time	20.08% (209)	15% (9)	
Paving street	Always present	26.22% (273)	26.67% (16)	1
	Absent in any time	73.78% (768)	73.33% (44)	
Sewage	Always present	45.24% (471)	31.67% (19)	0.054
	Absent in any time	54.76% (570)	68.33% (41)	
Helminthic infection	Yes	27.19% (283)	20% (12)	0.284
	No	72.81% (758)	80% (48)	

Table S4. IL-13 production in SCAALA cohort according to sociodemographic and environmental variables.

Variable		Non-producers	Producers	p-value
Sex	Female	45.72% (347)	46.87% (187)	0.756
	Male	54.28% (412)	53.13% (212)	
	Age	6 (5-8)	6 (5-8)	0.455
	BMI Z-score	-0.30 (-0.96 - 0.44)	-0.27 (-0.92 - 0.56)	0.365
Birth weight	≥ 2500g	91.6% (687)	90.89% (359)	0.766
	< 2500g	8.4% (63)	9.11% (36)	
Infection Burden	Light	40.93% (307)	42.03% (166)	0.769
	Heavy	59.07% (443)	57.97% (229)	
Mother education	Only primary school	22% (165)	21.27% (84)	0.868
	More than primary school	47.6% (357)	46.84% (185)	
	More than high school	30.4% (228)	31.9% (126)	
Income	< 1 MW	53.87% (404)	54.68% (216)	0.746
	1 - 2 MW	29.47% (221)	30.38% (120)	
	>2 MW	16.67% (125)	14.94% (59)	
Day care attendance	Yes	16.13% (121)	16.71% (66)	0.868
	No	83.87% (629)	83.29% (329)	
Smoker in home	Yes	26.8% (201)	26.58% (105)	0.993
	No	73.2% (549)	73.42% (290)	
Mold in home	Yes	70.8% (531)	65.57% (259)	0.08
	No	29.2% (219)	34.43% (136)	
Tap water	Always present	80.13% (601)	79.24% (313)	0.779
	Absent in any time	19.87% (149)	20.76% (82)	
Paving street	Always present	26.93% (202)	22.78% (90)	0.144
	Absent in any time	73.07% (548)	77.22% (305)	
Sewage	Always present	44.67% (335)	43.29% (171)	0.702
	Absent in any time	55.33% (415)	56.71% (224)	
Helminthic infection	Yes	24.93% (187)	30.13% (119)	0.069
	No	75.07% (563)	69.87% (276)	

Table S5. IL-10 production in SCAALA cohort according to sociodemographic and environmental variables.

Variable		Non-producers	Producers	p-value
Sex	Female	46.58% (517)	41.18% (42)	0.346
	Male	53.42% (593)	58.82% (60)	
	Age	6 (5-8)	6 (5-8)	0.149
	BMI Z-score	-0.29 (-0.96 - 0.52)	-0.28 (-1.11 - 0.33)	0.386
Birth weight	≥ 2500g	90.98% (999)	96.04% (97)	0.121
	< 2500g	9.02% (99)	3.96% (4)	
Infection Burden	Light	59.47% (653)	47.52% (48)	0.026
	Heavy	40.53% (445)	52.48% (53)	
Mother education	Only primary school	21.58% (237)	26.73% (27)	0.43
	More than primary school	47.36% (520)	46.53% (47)	
	More than high school	31.06% (341)	26.73% (27)	
Income	< 1 MW	54.1% (594)	57.43% (58)	0.809
	1 - 2 MW	29.51% (324)	27.72% (28)	
	>2 MW	16.39% (180)	14.85% (15)	
Day care attendance	Yes	16.3% (179)	12.87% (13)	0.448
	No	83.7% (919)	87.13% (88)	
Smoker in home	Yes	27.05% (297)	23.76% (24)	0.551
	No	72.95% (801)	76.24% (77)	
Mold in home	Yes	69.31% (761)	66.34% (67)	0.613
	No	30.69% (337)	33.66% (34)	
Tap water	Always present	80.51% (884)	74.26% (75)	0.17
	Absent in any time	19.49% (214)	25.74% (26)	
Paving street	Always present	26.5% (291)	7.92% (8)	<0.001
	Absent in any time	73.5% (807)	92.08% (93)	
Sewage	Always present	46.63% (512)	15.84% (16)	<0.001
	Absent in any time	53.37% (586)	84.16% (85)	
Helminthic infection	Yes	25.59% (281)	42.57% (43)	<0.001
	No	74.41% (817)	57.43% (58)	

Table S6. Admixture Mapping statistics of identified suggestive peaks for IL-13 production among children from the Social Change, Asthma, Allergy in Latin America (SCAALA) Cohort in Salvador, Brazil. Those suggestive peaks were identified in Table 2. and windows associated were investigated for the other ancestries.

Ancestry	Chr Region	Chr Position	Initial Window Marker	Final Window Marker	Effect (OR)	P
African	11q14.1	11:84380621 - 84612058	rs17147488	rs7943571	0.78 (0.65 - 0.94)	1.05E-02
	16q23.3	16:82828365 - 82865785	'	rs9936998	1.32 (1.09 - 1.59)	4.33E-03
	17p13.2	17:3564235 - 3672941	rs3826493	rs11649949	1.28 (1.06 - 1.54)	1.18E-02
		17:3674153 - 3703782	rs12604024	rs4300684	1.27 (1.05 - 1.53)	1.35E-02
		17:3705526 - 3762271	rs2891	rs57724104	1.26 (1.04 - 1.53)	1.63E-02
		17:3763574 - 3794352	rs1558294	rs7216488	1.32 (1.09 - 1.6)	4.21E-03
17:4455500 - 4490468	rs7206974	rs7210765	1.23 (1.02 - 1.48)	3.17E-02		
Native American	11q14.1	11:84380621 - 84612058	rs17147488	rs7943571	1.34 (0.95 - 1.9)	9.96E-02
	16q23.3	16:82828365 - 82865785	rs12929663	rs9936998	1.20 (0.84 - 1.7)	3.15E-01
	17p13.2	17:3564235 - 3672941	rs3826493	rs11649949	0.71 (0.48 - 1.04)	7.60E-02
		17:3674153 - 3703782	rs12604024	rs4300684	0.71 (0.48 - 1.03)	7.35E-02
		17:3705526 - 3762271	rs2891	rs57724104	0.70 (0.48 - 1.02)	6.58E-02
		17:3763574 - 3794352	rs1558294	rs7216488	0.80 (0.56 - 1.16)	2.45E-01
17:4455500 - 4490468	rs7206974	rs7210765	0.65 (0.44 - 0.94)	2.30E-02		
African	3p21.2 - 3p21.1	3:51368795 - 51962081	rs78870156	rs323881	0.95 (0.78 - 1.14)	5.67E-01
		3:51963904 - 52536308	rs11921171	rs758801	0.95 (0.79 - 1.15)	5.93E-01
		3:52544155 - 53178658	rs114506140	rs62254273	0.94 (0.78 - 1.14)	5.50E-01
	10q21.3	10:68276614 - 68396283	rs76422529	rs7072222	1.18 (0.98 - 1.41)	8.81E-02
	15q25.1	15:80563203 - 80637884	rs11856226	rs9920521	1.02 (0.84 - 1.24)	8.23E-01
		15:80639564 - 80754984	rs3848198	rs8028295	1.01 (0.83 - 1.22)	9.32E-01
		15:80756302 - 80878680	rs115471347	rs73494771	1.02 (0.84 - 1.23)	8.75E-01
		15:80879709 - 81077104	rs76810611	rs12372942	1.02 (0.84 - 1.23)	8.37E-01
15:81077470 - 81104785	rs10519292	rs28468684	1.05 (0.87 - 1.27)	6.35E-01		
European	3p21.2 - 3p21.1	3:51368795 - 51962081	rs78870156	rs323881	1.12 (0.92 - 1.35)	2.56E-01
		3:51963904 - 52536308	rs11921171	rs758801	1.13 (0.93 - 1.37)	2.20E-01
		3:52544155 - 53178658	rs114506140	rs62254273	1.12 (0.92 - 1.35)	2.64E-01
	10q21.3	10:68276614 - 68396283	rs76422529	rs7072222	0.99 (0.82 - 1.19)	9.22E-01
	15q25.1	15:80563203 - 80637884	rs11856226	rs9920521	0.86 (0.71 - 1.04)	1.19E-01
		15:80639564 - 80754984	rs3848198	rs8028295	0.84 (0.7 - 1.02)	8.51E-02
		15:80756302 - 80878680	rs115471347	rs73494771	0.85 (0.7 - 1.03)	1.06E-01
		15:80879709 - 81077104	rs76810611	rs12372942	0.86 (0.71 - 1.04)	1.25E-01
15:81077470 - 81104785	rs10519292	rs28468684	0.88 (0.73 - 1.07)	2.06E-01		

Table S7. Local ancestry window in 16q23.3 associations for IL-13 adjusted for previously associated variants. Analysis obtained using logistic regression including ancestries windows (16:82828365 – 82865785), SNVs and covariates in the model.

Local Ancestry	OR (CI 95%)	P-value	SNV	OR (CI 95%)	P-value
African	1.20 (0.98 - 1.46)	8.17E-02	rs12598402	0.72 (0.57 - 0.92)	8.02E-03
	1.23 (1.01 - 1.50)	3.65E-02	rs28376615	1.57 (1.13 - 2.18)	7.27E-03
	1.23 (1.01 - 1.50)	3.63E-02	rs75929636	1.49 (1.11 - 2.00)	7.39E-03
European	0.73 (0.58 - 0.93)	1.10E-02	rs12598402	0.77 (0.64 - 0.94)	8.75E-03
	0.77 (0.63 - 0.93)	7.98E-03	rs28376615	1.55 (1.12 - 2.15)	8.61E-03
	0.77 (0.64 - 0.94)	8.75E-03	rs75929636	1.47 (1.1 - 1.97)	9.71E-03
Native American	1.23 (0.87 - 1.75)	2.49E-01	rs12598402	0.67 (0.53 - 0.84)	4.86E-04
	1.23 (0.86 - 1.74)	2.53E-01	rs28376615	1.73 (1.26 - 2.38)	7.26E-04
	1.21 (0.85 - 1.72)	2.53E-01	rs75929636	1.62 (1.22 - 2.16)	8.22E-04

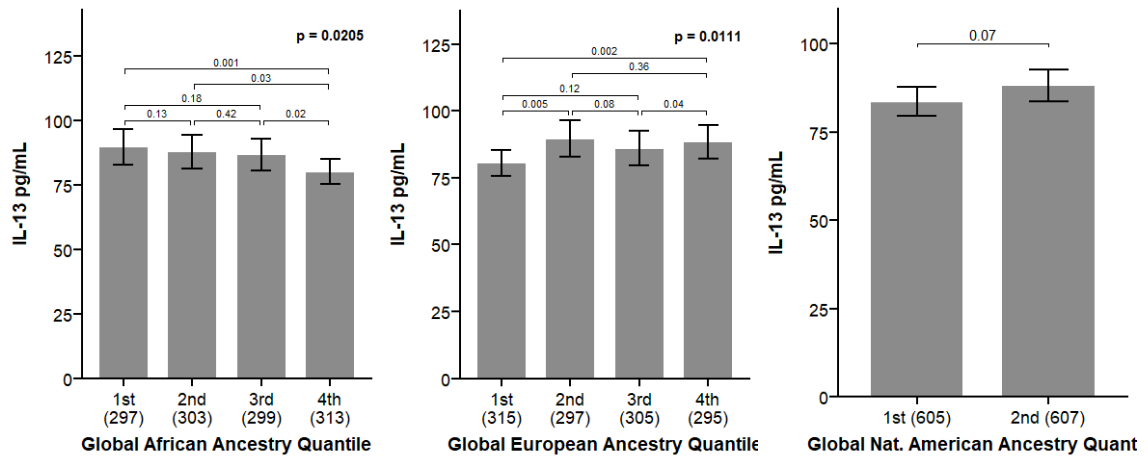


Figure S1. Global ancestry quantiles comparison for IL-13 concentration. Concentrations of IL-13 are expressed in geometric mean and 95% confidence interval. The p-values were from the Kruskal-Wallis test with Dunn post-test (when necessary).

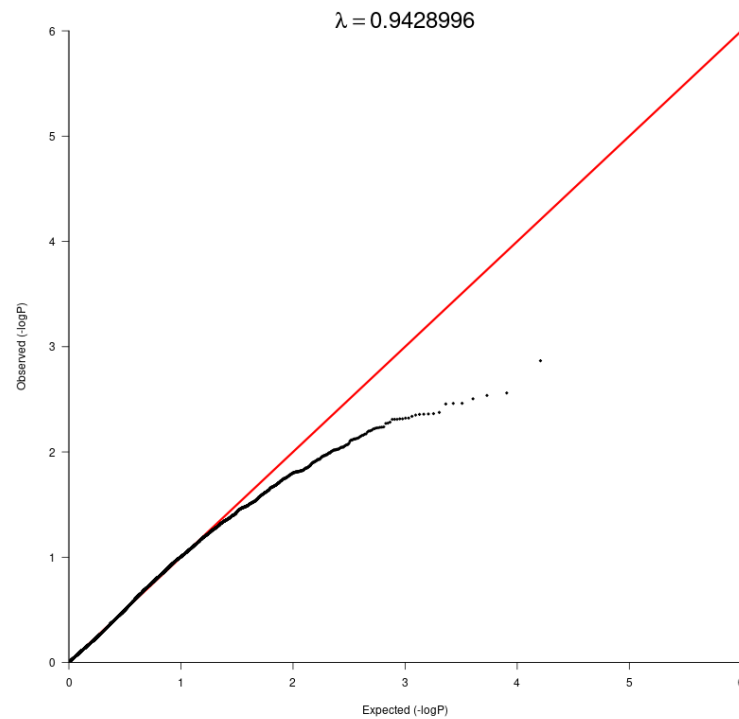


Figure S2. Quantile-quantile (QQ) plot for IL-13 production association with African Local Ancestry. P-values correspond to the data shown in the Manhattan plot (Figure 1A).

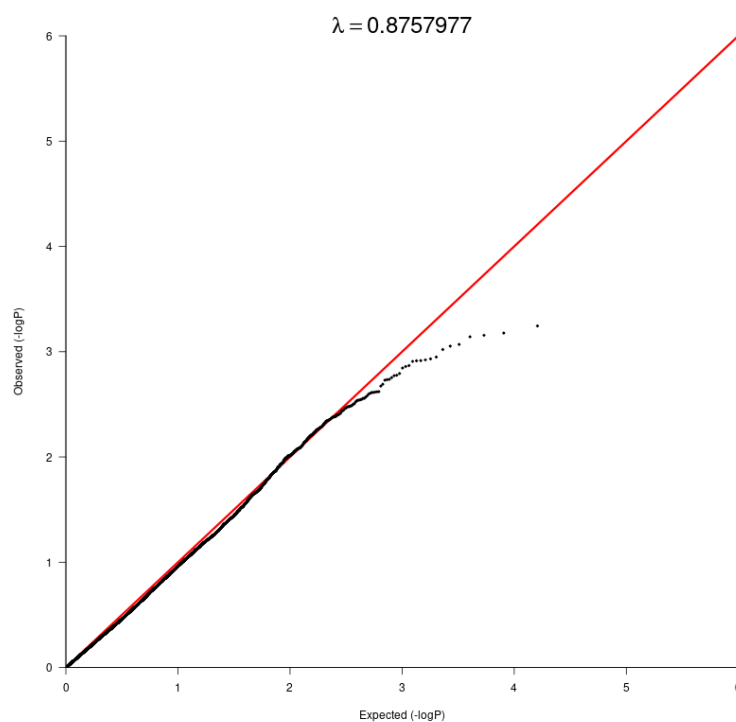


Figure S3. Quantile-quantile (QQ) plot for IL-13 production association with European Local Ancestry. P-values correspond to the data shown in the Manhattan plot (Figure 1B).

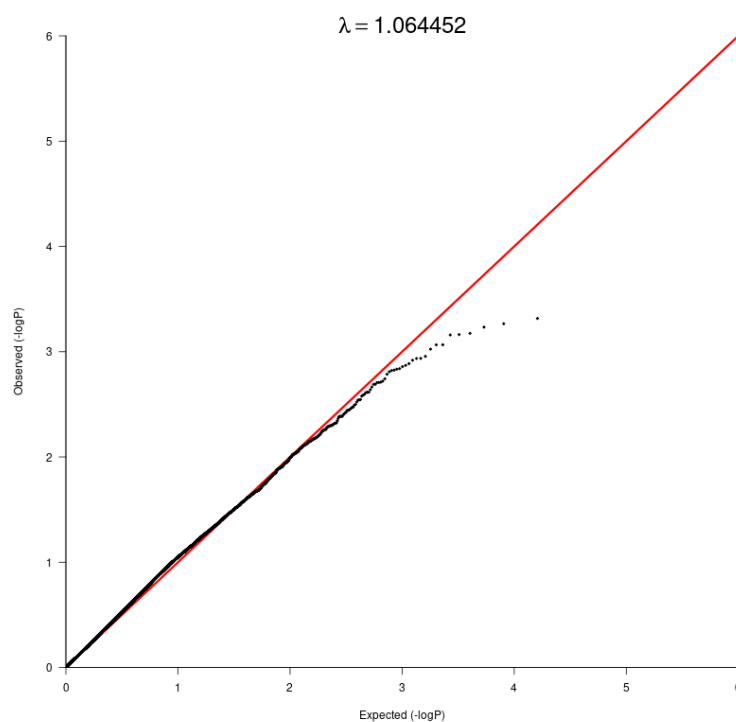


Figure S4. Quantile-quantile (QQ) plot for IL-13 production association with Native American Local Ancestry. P-values corresponding to the data shown in the Manhattan plot (Figure 1C).

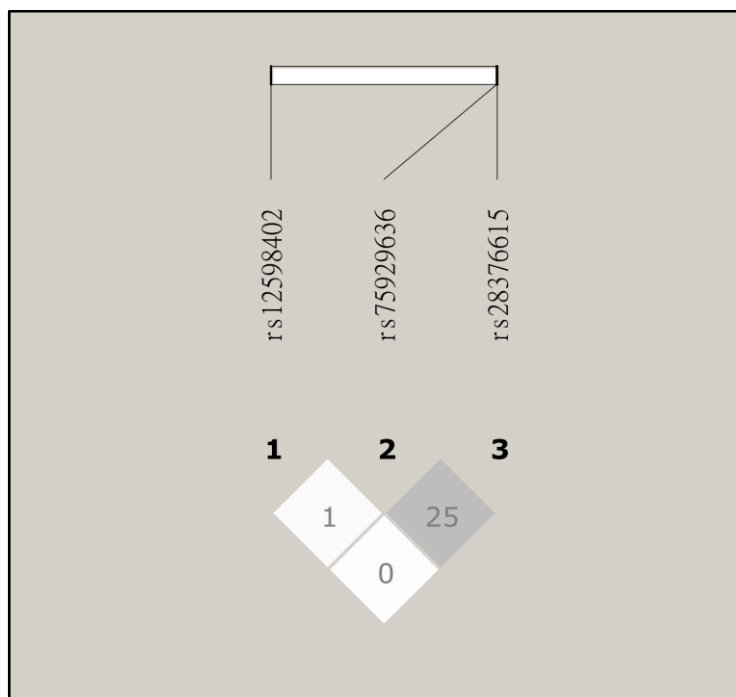


Figure S5. Linkage Disequilibrium (LD) Plot for associated variants rs12598402, rs28376615 and rs75929636. LD calculated by using HaploView software. LD plot showing pairwise r -squared values.

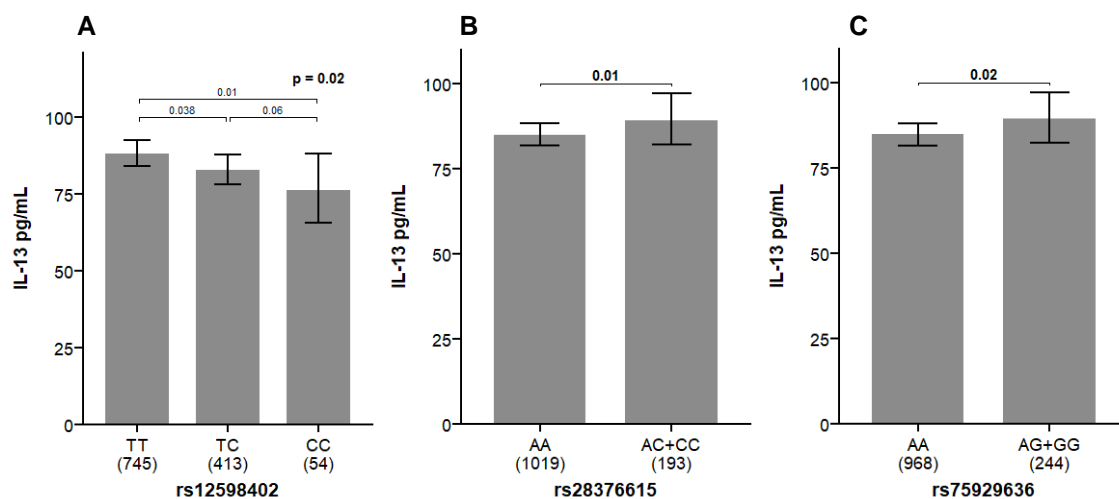


Figure S6. IL-13 concentration and rs12598402 (A), rs28376615 (B) and rs75929636 (C) genotype. Concentrations of IL-13 are expressed in geometric mean and 95% confidence interval. The p-values were from the Kruskal-Wallis test with Dunn post-test (when necessary).

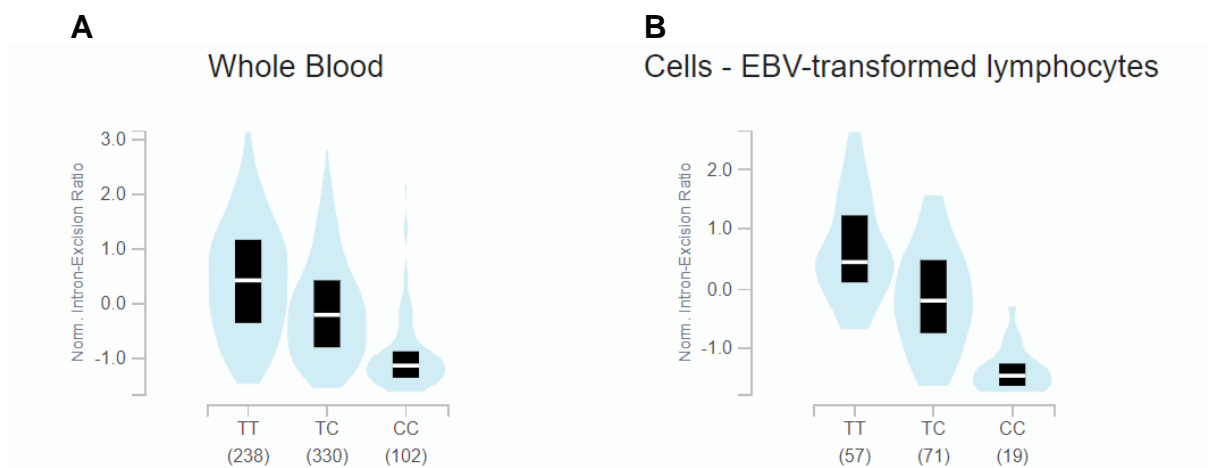


Figure S7. Violin plot of *PLCG2* intron-excision ratios according to rs12598402 genotypes in Genotype-Tissue Expression (GTEx). The allelic effect of rs34767465 as splicing quantitative trait locus (sQTL) in Whole Blood (A) and EBV-transformed lymphocytes (B). Normalized *PLCG2* gene intron-excision ratios are shown by box-plots within violin plots. Minor allele normalized effect size was -0.63 ($p = 2.1 \times 10^{-30}$) and -0.80 (1.4×10^{-12}) for Whole Blood (A) and EBV-transformed lymphocytes (B), respectively.

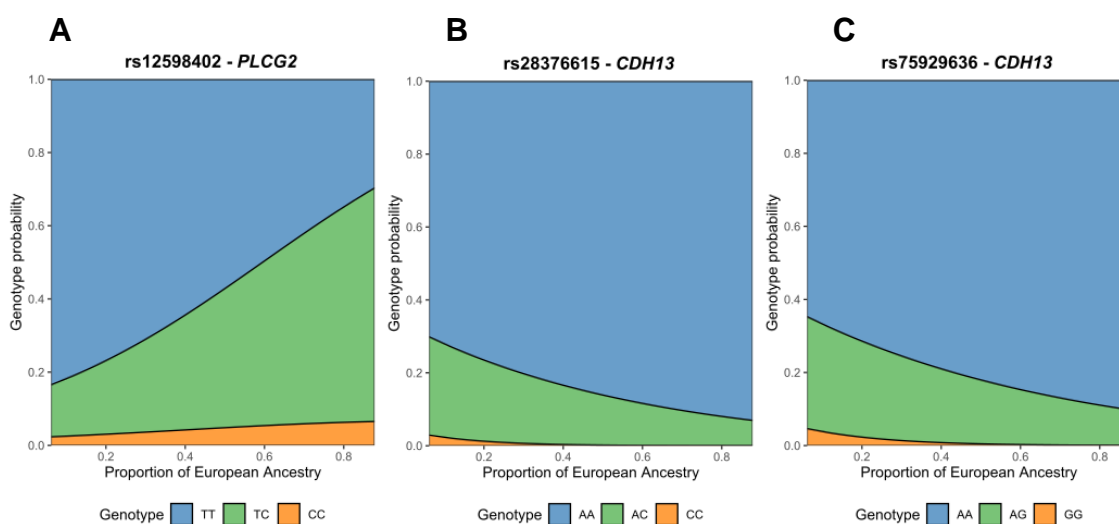


Figure S8. Frequency distribution of genotypes according to the estimated European genomic ancestry for the SCAALA population. Fitted multinomial logistic regression models showing the association between the distribution of genotypes for (A) rs12598402, (B) 28376615 and (C) 75929636 according to European ancestry.

DISCUSSÃO

O sistema imunológico é um forte alvo de pressão seletiva, especialmente considerando que ao longo da história, os humanos compartilharam o ambiente e conviveram com diversos organismos, incluindo microrganismos e parasitas multicelulares. Desta forma, é esperado que a composição genética dos humanos atuais relacionada ao sistema imune seja produto de diversos processos evolutivos, incluindo a seleção de variantes genéticas que permitam conviver com esses outros seres. Nessa perspectiva, considerando aspectos da evolução e migração humana, concomitante a mudanças ambientais, foi investigada a influência da ancestralidade genética individual na produção das citocinas IFN- γ , IL-5, IL-13 e IL-10, que estão relacionadas a contextos distintos da resposta imune em uma população brasileira.

O aumento na proporção de ancestralidade genômica Africana foi relacionado a uma menor chance de produção da citocina IL-13, envolvida na resposta anti-helmíntica que é do tipo Th2. Ademais, a ancestralidade genômica Europeia, que está inversamente correlacionada à Africana na população estudada, apresentou a associação no sentido inverso. Esse achado traz luz na compreensão de como a ancestralidade pode influenciar no viés para um dado tipo de resposta. Vale ressaltar que na literatura é trazido que um *background* genético mais relacionado a populações africanas está associado a uma maior resposta Th2 em um contexto livre de helmintíases. Nesse sentido, é importante responder essa pergunta no contexto do Brasil, pois o *background* genético das populações brasileiras é um produto da mistura de três populações ancestrais, os povos Nativo-Americanos, os Europeus e os Africanos. Além disso, trata-se de um país em desenvolvimento cuja presença de infecções helmínticas ainda é um aspecto encontrado em algumas regiões do país. A exposição a longo prazo a tais parasitas pode induzir mecanismos regulatórios e imunomodulatórios no hospedeiro, relacionados à produção da citocina IL-10, por exemplo.

Dado que a IL-13 se mostrou associada com a ancestralidade, um outro aspecto investigado no presente estudo foi a relação com a ancestralidade local, ou seja, a origem biogeográfica de segmentos cromossômicos ao longo do genoma. Essa abordagem permite investigar regiões específicas cuja ancestralidade pode estar associada ao fenótipo, e com posterior análise mais refinada da região identificar

variantes genéticas associadas. Portanto, a partir do mapeamento por miscigenação e subsequente mapeamento fino da região identificamos três SNVs associados com a produção de IL-13 em dois genes do cromossomo 16, cujo os achados foram replicados em uma outra coorte. Ademais, os alelos destas variantes genéticas associados a uma maior chance de produzir IL-13 apresentam uma maior probabilidade de serem encontrados em populações de origem africana. Desta forma, o achado está de acordo com o padrão esperado do ponto de vista genético-populacional, em que variantes genéticas associadas a uma resposta Th2 acentuada estão enriquecidas no *background* Africano.

CONCLUSÕES

Em conclusão, encontramos uma associação entre ancestralidade genômica biogeográfica e produção da citocina de assinatura Th2 IL-13, o que levou a realizar três mapeamentos de miscigenação e posterior mapeamento fino para produção de IL-13. A partir disso identificamos três SNVs enriquecidos por ancestralidade que estão associados à produção de IL-13, nos genes *PLCG2* e *CDH13*, em que a variante identificada no gene *PLCG2* está potencialmente relacionada a mudanças no perfil de expressão em tecido sanguíneo e células imune. Nossos resultados mostram que, de fato, variantes genéticas ligadas a uma maior chance de produção de IL-13 são enriquecidas em populações de ascendência africana e também podem afetar a produção de IL-13 em uma população miscigenada. Estudos futuros podem um painel ainda maior de citocinas e outros mediadores, além também da frequência de células imunes. Investigações nessa perspectiva podem contribuir para melhor compreensão da variação imune em nível populacional no contexto do Brasil.

REFERÊNCIAS

1. BARREIRO, Luis B.; QUINTANA-MURCI, Lluís. Evolutionary and population (epi) genetics of immunity to infection. **Human genetics**, v. 139, n. 6, p. 723-732, 2020.
2. BEDOYA, Simone Kennedy *et al.* Th17 cells in immunity and autoimmunity. **Clinical and Developmental Immunology**, v. 2013, 2013.
3. BENTON, Mary Lauren *et al.* The influence of evolutionary history on human health and disease. **Nature Reviews Genetics**, v. 22, n. 5, p. 269-283, 2021.
4. BRODIN, Petter *et al.* Variation in the human immune system is largely driven by non-heritable influences. **Cell**, v. 160, n. 1-2, p. 37-47, 2015.
5. BRODIN, Petter; DAVIS, Mark M. Human immune system variation. **Nature reviews immunology**, v. 17, n. 1, p. 21-29, 2017.
6. BUSZKO, Maja; SHEVACH, Ethan M. Control of regulatory T cell homeostasis. **Current Opinion in Immunology**, v. 67, p. 18-26, 2020.
7. CARR, Edward J. *et al.* The cellular composition of the human immune system is shaped by age and cohabitation. **Nature immunology**, v. 17, n. 4, p. 461-468, 2016.
8. CRAIG, Jeffrey M. *et al.* The value of twins for health and medical research: a third of a century of Progress. **Twin Research and Human Genetics**, v. 23, n. 1, p. 8-15, 2020.
9. CRESPI, Bernard J.; GO, Matthew C. Diametrical diseases reflect evolutionary-genetic tradeoffs: evidence from psychiatry, neurology, rheumatology, oncology and immunology. **Evolution, Medicine, and Public Health**, v. 2015, n. 1, p. 216-253, 2015.
10. DOMÍNGUEZ-ANDRÉS, Jorge; NETEA, Mihai G. Impact of historic migrations and evolutionary processes on human immunity. **Trends in Immunology**, v. 40, n. 12, p. 1105-1119, 2019.
11. DUPAGE, Michel; BLUESTONE, Jeffrey A. Harnessing the plasticity of CD4+ T cells to treat immune-mediated disease. **Nature Reviews Immunology**, v. 16, n. 3, p. 149-163, 2016.
12. EIZENBERG-MAGAR, Inbal *et al.* Diverse continuum of CD4+ T-cell states is determined by hierarchical additive integration of cytokine signals. **Proceedings of the National Academy of Sciences**, v. 114, n. 31, p. E6447-E6456, 2017.
13. FREW, John W. The hygiene hypothesis, old friends, and new genes. **Frontiers in Immunology**, v. 10, p. 388, 2019.
14. FUMAGALLI, Matteo *et al.* The landscape of human genes involved in the immune response to parasitic worms. **BMC evolutionary biology**, v. 10, n. 1, p. 1-15, 2010.

15. FUMAGALLI, Matteo *et al.* Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. **PLoS genetics**, v. 7, n. 11, p. e1002355, 2011.
16. GARN, Holger; POTACZEK, Daniel Piotr; PFEFFERLE, Petra Ina. The hygiene hypothesis and new perspectives—Current challenges meeting an old postulate. **Frontiers in Immunology**, v. 12, p. 847, 2021.
17. HAAHTELA, Tari. A biodiversity hypothesis. **Allergy**, v. 74, n. 8, p. 1445-1456, 2019.
18. IDAGHDOUR, Youssef *et al.* Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nature genetics*, v. 42, n. 1, p. 62-67, 2010.
19. IWASAKI, Akiko; MEDZHITOV, Ruslan. Control of adaptive immunity by the innate immune system. **Nature immunology**, v. 16, n. 4, p. 343-353, 2015.
20. KANY, Shinwan; VOLLRATH, Jan Tilmann; RELJA, Borna. Cytokines in inflammatory disease. **International journal of molecular sciences**, v. 20, n. 23, p. 6008, 2019.
21. KARLSSON, Elinor K.; KWIATKOWSKI, Dominic P.; SABETI, Pardis C. Natural selection and infectious disease in human populations. **Nature Reviews Genetics**, v. 15, n. 6, p. 379-393, 2014.
22. KOLLMANN, Tobias R. Variation between populations in the innate immune response to vaccine adjuvants. **Frontiers in immunology**, v. 4, p. 81, 2013.
23. KORT, Remco *et al.* Shaping the oral microbiota through intimate kissing. **Microbiome**, v. 2, n. 1, p. 1-8, 2014.
24. LAX, Simon *et al.* Longitudinal analysis of microbial interaction between humans and the indoor environment. **Science**, v. 345, n. 6200, p. 1048-1052, 2014.
25. LEE, Jongdae *et al.* The multifaceted role of Th1, Th9, and Th17 cells in immune checkpoint inhibition therapy. **Frontiers in Immunology**, v. 12, 2021.
26. LISTON, Adrian; CARR, Edward J.; LINTERMAN, Michelle A. Shaping variation in the human immune system. **Trends in immunology**, v. 37, n. 10, p. 637-646, 2016.
27. LISTON, Adrian *et al.* Human immune diversity: from evolution to modernity. **Nature immunology**, v. 22, n. 12, p. 1479-1489, 2021.
28. LIU, Chao *et al.* Cytokines: From clinical significance to quantification. **Advanced Science**, v. 8, n. 15, p. 2004433, 2021.
29. LUCKHEERAM, Rishi Vishal *et al.* CD4+ T cells: differentiation and functions. **Clinical and developmental immunology**, v. 2012, 2012.

30. MANGINO, Massimo *et al.* Innate and adaptive immune traits are differentially affected by genetic and environmental factors. **Nature communications**, v. 8, n. 1, p. 1-7, 2017.
31. MARSHALL, Jean S. *et al.* An introduction to immunology and immunopathology. **Allergy, Asthma & Clinical Immunology**, v. 14, n. 2, p. 1-10, 2018.
32. MARTINEZ-SANCHEZ, Mariana E. *et al.* Role of cytokine combinations on CD4+ T cell differentiation, partial polarization, and plasticity: continuous network modeling approach. **Frontiers in physiology**, v. 9, p. 877, 2018.
33. NÉDÉLEC, Yohann, *et al.* Genetic ancestry and natural selection drive population differences in immune responses to pathogens. **Cell**, 2016, 167.3: 657-669. e21.
34. NIELSEN, Rasmus *et al.* Tracing the peopling of the world through genomics. **Nature**, v. 541, n. 7637, p. 302-310, 2017.
35. NORRIS, Emily T. *et al.* Genetic ancestry, admixture and health determinants in Latin America. **BMC genomics**, v. 19, n. 8, p. 75-87, 2018.
36. OKADA, Hea *et al.* The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. **Clinical & Experimental Immunology**, v. 160, n. 1, p. 1-9, 2010.
37. ORRÙ, Valeria *et al.* Genetic variants regulating immune cell levels in health and disease. **Cell**, v. 155, n. 1, p. 242-256, 2013.
38. PAWLAK, Mathias; HO, Allen W.; KUCHROO, Vijay K. Cytokines and transcription factors in the differentiation of CD4+ T helper cell subsets and induction of tissue inflammation and autoimmunity. **Current opinion in immunology**, v. 67, p. 57-67, 2020.
39. PFEFFERLE, Petra I. *et al.* The Hygiene Hypothesis—Learning From but Not Living in the Past. **Frontiers in Immunology**, v. 12, p. 632, 2021.
40. PENG, Kerui *et al.* Diversity in immunogenomics: the value and the challenge. **Nature Methods**, v. 18, n. 6, p. 588-591, 2021.
41. RAPHAEL, Itay *et al.* T cell subsets and their signature cytokines in autoimmune and inflammatory diseases. **Cytokine**, v. 74, n. 1, p. 5-17, 2015.
42. QUACH, Hélène *et al.* Genetic adaptation and Neandertal admixture shaped the immune system of human populations. **Cell**, v. 167, n. 3, p. 643-656. e17, 2016.
43. ROMAGNANI, Sergio. The increased prevalence of allergy and the hygiene hypothesis: missing immune deviation, reduced immune suppression, or both?. **Immunology**, v. 112, n. 3, p. 352-363, 2004.
44. ROTHSCCHILD, Daphna *et al.* Environment dominates over host genetics in shaping human gut microbiota. **Nature**, v. 555, n. 7695, p. 210-215, 2018.

45. ROOK, Graham AW; MARTINELLI, Roberta; BRUNET, Laura Rosa. Innate immune responses to mycobacteria and the downregulation of atopic responses. **Current opinion in allergy and clinical immunology**, v. 3, n. 5, p. 337-342, 2003.
46. ROOK, G. A. W.; BRUNET, L. R. Microbes, immunoregulation, and the gut. **Gut**, v. 54, n. 3, p. 317-320, 2005.
47. RUSSELL, Shirley B. *et al.* Pleiotropic effects of immune responses explain variation in the prevalence of fibroproliferative diseases. **PLoS genetics**, v. 11, n. 11, p. e1005568, 2015.
48. SAHU, Monalisha; PRASUNA, Josyula G. Twin studies: A unique epidemiological tool. **Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine**, v. 41, n. 3, p. 177, 2016.
49. SANZ, Joaquin; RANDOLPH, Haley E.; BARREIRO, Luis B. Genetic and evolutionary determinants of human population variation in immune responses. **Current Opinion in Genetics & Development**, v. 53, p. 28-35, 2018.
50. SCUDELLARI, Megan. News Feature: Cleaning up the hygiene hypothesis. **Proceedings of the National Academy of Sciences**, v. 114, n. 7, p. 1433-1436, 2017.
51. SHEN-ORR, Shai S. *et al.* Defective signaling in the JAK-STAT pathway tracks with chronic inflammation and cardiovascular risk in aging humans. **Cell systems**, v. 3, n. 4, p. 374-384. e4, 2016.
52. SIMON, A. Katharina; HOLLANDER, Georg A.; MCMICHAEL, Andrew. Evolution of the immune system in humans from infancy to old age. **Proceedings of the Royal Society B: Biological Sciences**, v. 282, n. 1821, p. 20143085, 2015.
53. SIRONI, Manuela; CLERICI, Mario. The hygiene hypothesis: an evolutionary perspective. **Microbes and Infection**, v. 12, n. 6, p. 421-427, 2010.
54. SONG, Se Jin *et al.* Cohabiting family members share microbiota with one another and with their dogs. **elife**, v. 2, p. e00458, 2013.
55. STRACHAN, David P. Hay fever, hygiene, and household size. **BMJ: British Medical Journal**, v. 299, n. 6710, p. 1259, 1989.
56. SUAREZ-PAJES, Eva *et al.* Genetic ancestry inference and its application for the genetic mapping of human diseases. **International Journal of Molecular Sciences**, v. 22, n. 13, p. 6962, 2021.
57. TER HORST, Rob *et al.* Host and environmental factors influencing individual human cytokine responses. **Cell**, v. 167, n. 4, p. 1111-1124. e13, 2016.
58. TSANG, John S. *et al.* Global analyses of human immune variation reveal baseline predictors of postvaccination responses. **Cell**, v. 157, n. 2, p. 499-513, 2014.

59. VAN DEN BIGGELAAR, Anita HJ *et al.* Decreased atopy in children infected with *Schistosoma haematobium*: a role for parasite-induced interleukin-10. **The Lancet**, v. 356, n. 9243, p. 1723-1727, 2000.
60. VAN DONGEN, Jenny *et al.* The continuing value of twin studies in the omics era. **Nature Reviews Genetics**, v. 13, n. 9, p. 640-653, 2012.
61. VATSIOU, Alexandra I.; BAZIN, Eric; GAGGIOTTI, Oscar E. Changes in selective pressures associated with human population expansion may explain metabolic and immune related pathways enriched for signatures of positive selection. **BMC genomics**, v. 17, n. 1, p. 1-11, 2016.
62. WILFING, Astrid *et al.* African-European differences in the capacity of T-cell cytokine production. **The American journal of tropical medicine and hygiene**, v. 65, n. 5, p. 504-509, 2001.
63. YAO, Song, *et al.* Genetic ancestry and population differences in levels of inflammatory cytokines in women: role for evolutionary selection and environmental factors. **PLoS genetics**, 2018, 14.6: e1007368.
64. YE, Chun Jimmie *et al.* Intersection of population variation and autoimmunity genetics in human T cell activation. **Science**, v. 345, n. 6202, p. 1254665, 2014.
65. ZHAO, Jingyue; LLOYD, Clare M.; NOBLE, Alistair. Th17 responses in chronic allergic airway inflammation abrogate regulatory T-cell-mediated tolerance and contribute to airway remodeling. **Mucosal immunology**, v. 6, n. 2, p. 335-346, 2013.
66. ZHU, Xiaoliang; ZHU, Jinfang. CD4 T helper cell subsets and related human immunological disorders. **International Journal of Molecular Sciences**, 2020, 21.21: 8011.