

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA



TESE DE DOUTORADO

Epidemiologia Genética do Câncer Gástrico na População Peruana

Orientada: Latife Pereira Lacorte

Orientador: Prof. Eduardo M. Tarazona Santos

Belo Horizonte
Agosto – 2012

Latife Pereira Lacorte

Epidemiologia Genética do Câncer Gástrico na População Peruana

Tese apresentada ao Programa de Pós-Graduação em Genética do Departamento de Biologia Geral do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de doutor em Genética.

Orientador: Eduardo Martín Tarazona Santos

Belo Horizonte
Agosto – 2012

Lacorte, Latife Pereira.
Epidemiologia genética do câncer gástrico na população peruana.
[manuscrito] / Latife Pereira Lacorte. – 2012.
97 f. : il. ; 29,5 cm

Orientador: Eduardo Martín Tarazona Santos.

Tese (doutorado) – Universidade Federal de Minas Gerais. Departamento de Biologia Geral.

1. Epidemiologia - Teses. 2. Estômago - Câncer - Teses. 3. Marcadores genéticos - Teses. 4. Ancestralidade – Teses. 5. Peru – Condições sociais – Teses. 6. Índios - Teses. 7. Epidemiologia genética - Teses. I. Tarazona Santos, Eduardo Martín. II. Universidade Federal de Minas Gerais. Departamento de Biologia Geral. III. Título.

575.1:61



Pós-Graduação em Genética
Departamento de Biologia Geral, ICB
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 - C.P. 486 - Pampulha - 31270-901 - Belo Horizonte - MG
e-mail: pg-gen@icb.ufmg.br FAX: (+31) - 3409-2570



"Epidemiologia genética do Câncer Gástrico na População Peruana"

Latife Pereira Lacorte

Tese aprovada pela banca examinadora constituída pelos Professores:

Eduardo Martin Tarazona Santos - Orientador
UFMG

Regina Mingroni-Netto
USP/SP

Rinaldo Wellerson Pereira
UCB/DF

Dulciene Maria de Magalhães Queiroz
UFMG

Glaura da Conceição Franco
UFMG

Belo Horizonte, 08 de agosto de 2012.

***À minha querida mestra
Maria Cristina Lima de Castro***

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador Eduardo Tarazona por ter me ajudado a crescer como pessoa e como pesquisadora. Por acreditar em mim até mais que eu mesma...

Agradeço a minha família, que agora está mais completa com a minha pequena Fernanda. Em especial ao Gustavo pela paciência, amor e por toda dedicação, me ajudando em todas as dificuldades no doutorado e na vida. Ao meu pai, vó Dondoca, meus irmãos Stênio, Sérgio e Vitor e Tia Said pelo apoio e incentivo nas minhas decisões, vitórias e derrotas.

Ao pesquisador Robert Gilman, sua equipe e amigos da Universidad Peruana Cayetano Heredia. Agradeço aos meus amigos do laboratório LDGH, Camilão, Hanaisa, Rennan, Maíra, Silvana, Marília, Ferdi..., em especial à Moara, Roxana e Giordano por terem me ajudado muito... desde o início... e do laboratório LBEM, Daniela Lacerda e Marilza.

Agradeço também aos amigos que ganhei nesses anos de ICB, durante as disciplinas, seminários, colaborações, butecos e conversas de corredor.

Agradeço à instituição UFMG e a todos os professores pela contribuição na minha formação acadêmica.

E finalmente, para todos aqueles que contribuíram para este trabalho e ficaram de fora desta lista de agradecimentos, peço desculpas e meus sinceros agradecimentos.

1 Sumário

LISTA DE FIGURAS	6
LISTA DE ABREVIATURAS	9
RESUMO	10
ABSTRACT	11
PREFÁCIO	12
INTRODUÇÃO	13
1.1 Revisão Bibliográfica	13
Câncer Gástrico	13
Adenocarcinoma Gástrico	15
Estudos caso-controle: CGAS (Candidate Genes Association Studies) e GWAS (Genome Wide Association Studies)	16
Ancestralidade em estudos caso-controle	19
1.2 Objetivos	25
CAPÍTULO 1	26
SOCIOECONOMIC AND NUTRITIONAL FACTORS ACCOUNT FOR THE ASSOCIATION OF GASTRIC CANCER WITH AMERINDIAN ANCESTRY IN A LATIN AMERICAN ADMIXED POPULATION	27
CAPÍTULO 2	57
THE SINGLE NUCLEOTIDE POLYMORPHISMS IN THE <i>IL8</i> , <i>IL8RA</i> , <i>IL8RB</i> , <i>MSMB</i> , <i>FGFR2</i> AND <i>PTGS2</i> GENES AND RISK OF GASTRIC CANCER IN PERUVIAN POPULATION	58
CONCLUSÕES	83
REFERÊNCIAS BIBLIOGRÁFICAS	84
Influence of Arg72Pro polymorphisms of TP53 on the response of buccal cells to radiotherapy	89
Evolutionary dynamics of the human nadph oxidase genes CYBB, CYBA, NCF2 and NCF4: functional implications	91
ANEXOS	92
Mutational Analysis of Two Boys With the Severe Perinatally Lethal Melnick–Needles Syndrome	92
Currículo lattes documentado para as atividades dos últimos quatro anos	94

LISTA DE FIGURAS

Introdução

Figura 1. Progressão do tecido normal ao adenocarcinoma gástrico tipo intestinal.

Figura 2. Representação da miscigenação e do efeito da estratificação de populações em estudos de associação.

Figura 3. Representação do destino de uma região cromossômica de uma população miscigenada (daqui em diante denominada *cromossomo*), derivada de duas populações parentais (**A e B**), e questões relativas ao mapeamento por miscigenação por meio de estudos caso-controle (**C**).

Capítulo 1

Figure 1. Principal Component Analysis of our Peruvian samples of gastric cancer cases, their controls, and Native Americans in the context of HapMap-III European, African, and Mexican individuals, and distribution of Native American ancestry in cases and controls (box).

Figure S1. Barplot of individual ancestry estimated with the software Structure for Africans (red), Europeans (green), and Native Americans (blue), as well as gastric cancer cases and controls.

Figure S2. Scatterplot and Spearman correlation between individual Native American ancestry estimates by Structure versus Frappe (a) and Admixture (b) methods.

Capítulo 2

Figure 1. Pattern of LD of the genes: *IL8RB* in (A) Native American, (B) Cases, (C) Control; *PTGS2* in (D) African, (E) European, (F) Native American.

LISTA DE TABELAS

Introdução

Tabela 1. Estudos de associação de doenças com ancestralidade em populações Latino Americanas

Capítulo 1

Table 1. Socioeconomic, nutritional, and digestive-symptom-related variables and their association with gastric cancer and Native American ancestry.

Table S1. Distribution of cases and controls across hospitals and its association with Native American ancestry.

Table S2. Socioeconomic, nutritional, and digestive-symptom-related variables, their LOAD scores with the three first factors of the multivariate factor analysis and significance of Spearman correlation between individual values of the variables and coordinates on each factor.

Table S3. Classification of socioeconomic, nutritional, and digestive-symptom-related variables used in Table 1 and their values used in the statistical analysis.

Table S4. Allele frequencies in the populations included in this study for the 103 Ancestry Informative Markers used in the study.

Capítulo 2

Table 1. Intrapopulation diversity indexes and results of neutrality tests in the studied populations for *IL8*, *IL8RA*, *IL8RB* and *PTGS2* genes.

Table 2. Haplotype distribution of the 5'-UTR and promoter of *PTGS2* gene.

Table 3. Haplotype distribution of the 5'-UTR and promoter of *IL8* gene.

Table 4. Haplotype distribution of the 5'-UTR and promoter of *IL8RA* gene.

Table 5. Haplotype distribution of the 5'-UTR and promoter of *IL8RB* gene.

Table 6. Pair-wise F_{st} genetic distances between populations and the variance of D_{xy} (filled cells).

Table 7. Genotype frequencies (filled cells) and allelic frequencies in studied populations

Table 8. Genotype frequencies, p value and confidence interval of the regression analysis using an additive model.

Table S1. PCR conditions to amplify the promoter regions of *IL8*, *IL8RA*, *IL8RB* and *PTGS2* genes.

Table S2. Number of base pairs resequenced for each region in the *IL8*, *IL8RA*, *IL8RB* and *PTGS2* genes.

Table S3. Classification of variable and categories of the socioeconomic, nutritional and clinical information.

Table S4. Genotype frequencies from the *IL8*, *IL8RA*, *IL8RB* and *PTGS2* promoter and 5' UTR sequences for the three parental populations (African and European from Hapmap) and for case-control samples.

LISTA DE ABREVIATURAS

ADMIXMAP – Admixture mapping
AED – Adenocarcinoma tipo difuso
AEI – Adenocarcinoma tipo intestinal
AIMs – Marcadores Informativos de Ancestralidade
ANOVA – analysis of variance
Arg – Arginina
CGAS – Candidate Genes Association Studies
CVCD – Variantes comuns/Doenças comuns
CYBA – cytochrome b-245, alpha polypeptide
CYBB – cytochrome b-245, beta polypeptide
DL – Desequilíbrio de ligação
FGFR2 – Fibroblast Growth Factor Receptor 2
GWAS – Genome Wide Association Studies
HER2 – herstatin
HWE – Hardy-Weinberg equilibrium
IL1RN – interleukin 1 receptor antagonist
IL8 – interleukin 8
IL8RA – interleukin-8 receptor type A
IL8RB – interleukin-8 receptor type B
MAF – Alelo Menos Frequente
MSMB – β -microseminoprotein
NCF2 – neutrophil cytosolic factor 2
NCF4 – neutrophil cytosolic factor 4
OR – Odds ration
PCR – Polymerase chain reaction
PLCE1 – phospholipase C, epsilon 1
PMID - PubMed Identifier
Pro – Prolina
PTGS2 - prostaglandin-endoperoxide synthase 2
SNP – Single Nucleotide Polymorphisms
SPSS – Statistical Package for the Social Sciences
TERT – telomerase reverse transcriptase
TP53 – tumor protein p53

RESUMO

O câncer gástrico ocorre com alta incidência no Peru e é o tipo de câncer que causa o maior número de mortes. Essa alta incidência é parcialmente explicada por diferenças na prevalência de fatores de risco conhecidos. Fatores genéticos podem também contribuir para essa diferença na incidência. A população Peruana recebeu contribuição de três populações ancestrais: Africana, Européia e Nativo Americana. Nós usamos marcadores informativos de ancestralidade (AIMs) para estimar a ancestralidade de indivíduos Peruanos com e sem câncer gástrico para avaliar a associação entre câncer gástrico e ancestralidade controlando 43 variáveis sócioeconômicas, nutricionais e sintomas. Para sintetizar esse número de variáveis nós usamos análise fatorial multivariada. Nós genotipamos 106 AIMs em 241 casos de câncer gástrico e 300 controles sem câncer gástrico e calculamos as contribuições para cada indivíduo utilizando o programa Structure. O Odds ratio (OR) para ancestralidade, analisada como uma variável contínua, foi estimado usando regressão logística com idade e *scores* individuais de três fatores (resultantes da análise fatorial) incluídos como covariáveis. Um maior nível de ancestralidade Nativo Americana foi associado com um risco maior de desenvolver câncer gástrico. O OR calculado foi de 3,69 ($p= 0.011$), mas quando fatores de risco conhecidos foram incluídos a associação significativa com ancestralidade Nativo Americana não persistiu (OR= 1,28; $p= 0,69$). Nós também sequenciamos a região promotora de 4 genes em indivíduos Nativos Americanos e em uma amostra de casos e controles (25, 61 e 58 indivíduos respectivamente) e em seguida genotipamos seis polimorfismos: rs4073 (*IL8*), rs4674258 (*IL8RB*), rs689465 and rs689466 (*PTGS2*), rs10993994 (*MSMB*) and rs1219648 (*FGFR2*). O polimorfismo nos genes *MSMB* e *FGFR2* foram selecionados por já terem sido associados com outros tipos de cânceres em estudos de associação (Genome Wide ASsociation). A análise de regressão logística utilizando ancestralidade, idade e *scores* individuais da análise fatorial como covariáveis não foi significativa ($p>0,05$). Estudos posteriores utilizando uma amostragem maior e incluindo mais informações sobre os tumores seria interessante para determinar se essas variáveis poderiam ter influenciado em nosso estudo de associação.

ABSTRACT

Gastric cancer occurs with a high incidence in Peru and is the cancer that causes more number of deaths. This incidence is partially explained by differences in the prevalence of known risk factors. Genetic factors may also contribute to this difference in incidence. The Peruvian population received contribution of three ancestral populations: African, European and Native American. We used ancestry informative markers (AIMs) to estimate the ancestry of Peruvian gastric cancer cases and controls and assessed the association between gastric cancer and ancestry controlling 43 socioeconomic, nutritional variables and symptoms. In order to synthesize the number of variables we used multivariate factor analysis. We genotyped 106 AIMs in 241 gastric cancer cases and 300 controls. Odds ratios (OR) for ancestry modeled as a continuous variable was estimated using logistic regression with age and individual scores from the three factors included as covariates. Higher Native American ancestry was associated with increased gastric cancer risk. The OR was 3.69 ($p=0.011$), but when known risk factors were adjusted for, the association with Native American ancestry did not persist (OR= 1.28, $p=0.69$). We also sequenced promoter regions of the *IL8*, *IL8RA*, *IL8RB* and *PTGS2* genes in some Native American and case-control sampling (25, 61 and 58 individuals respectively) and after genotyped six polymorphisms: rs4073 (*IL8*), rs4674258 (*IL8RB*), rs689465 and rs689466 (*PTGS2*), rs10993994 (*MSMB*) and rs1219648 (*FGFR2*). The SNPs in the *MSMB* and *FGFR2* were selected because they were associated with other types of cancer in genome wide association studies. The p values from the logistic regression analysis using the ancestry (continuous variable), the age and the individual scores from the factorial analysis as covariates were not significant ($p>0.05$). Further work would be interesting with a larger sampling and including more information about the tumor to determine if these variables could have influenced in our association study.