

MÁRCIA HOLSBAACH BELTRAME

**ASPECTOS EVOLUTIVOS E IMPACTO FUNCIONAL DE
POLIMORFISMOS DOS GENES *CD80* E *CD86***

CURITIBA

2012

MÁRCIA HOLSBACH BELTRAME

**ASPECTOS EVOLUTIVOS E IMPACTO FUNCIONAL DE
POLIMORFISMOS DOS GENES *CD80* E *CD86***

Tese apresentada como requisito parcial à obtenção do título de Doutor. Programa de Pós Graduação em Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná.

Orientadora: Prof^a. Dr^a. Maria Luiza Petzl-Erler.

CURITIBA

2012



PARECER

Os abaixo-assinados, membros da Banca Examinadora da defesa de tese de Doutorado, a qual se submeteu **MÁRCIA HOLSBACH BELTRAME**, para obtenção do título de Doutora em Genética pela Universidade Federal do Paraná, no Programa de Pós-Graduação em Genética, são de parecer que se confira à candidata o conceito "A".

Secretaria da Coordenação do Programa de Pós-Graduação em Genética do Setor de Ciências Biológicas da Universidade Federal do Paraná.

Curitiba, 28 de fevereiro de 2012

Professor Doutor José Artur Bogo Chies
Universidade Federal do Rio Grande do Sul – Membro Titular

Professor Doutor Diogo Meyer
Universidade de São Paulo – Membro Titular

Professora Doutora Angélica Beate Winter Boldt
Dep. Patologia Médica/UFPR – Membro Titular

Professora Doutora Enilze Maria de Souza Fonseca Ribeiro
Dep. Genética/UFPR – Membro Titular

Professora Doutora Ana Claudia Bonatto
Dep. Genética/UFPR – Membro Titular

Professora Doutora Maria Luiza Petzi-Erler
Dep. Genética/UFPR – Orientadora e Presidente da Banca

Visto

Professora Doutora Enilze Maria de Souza Fonseca Ribeiro
Vice-Coordenadora do Programa de Pós-Graduação em Genética

AGRADECIMENTOS

À professora Luiza, pelos oito anos em que me orientou com muita competência, dedicação, amizade e carinho. Pelo cuidado que dedica a cada um de seus alunos e a seus projetos de pesquisa. Seu incentivo me fez ir além do que eu imaginava possível e sua postura científica é o exemplo que eu tento seguir.

Ao professor Arthur, que me recebeu de maneira tão acolhedora em seu laboratório e com muita dedicação me orientou durante o ano em que fiz meu estágio na Universidade de Yale.

À professora Roseli, que abriu as portas de seu laboratório e me orientou em todas as etapas da clonagem e do sequenciamento com o mesmo cuidado e critério que dedica aos seus alunos.

Ao professor Gabardo, que me orientou nas análises estatísticas não só realizando as análises, mas como excelente professor que é, me fazendo aprender todas as etapas. Por me presentear sempre com sua alegria, sua amizade e seus conselhos.

Aos meus professores do departamento de genética por todos os ensinamentos, sempre transmitidos com muita paixão.

À todos que são ou já foram meus colegas de laboratório, no LGMH ou no laboratório do professor Arthur, por me transmitirem todos os seus conhecimentos técnicos e científicos e por compartilharem comigo o dia-a-dia com suas conquistas e dificuldades, no laboratório e na vida. Especialmente à KK pelas clonagens realizadas em São Paulo e ao Ricardo pelas reações de sequenciamento enquanto eu estava nos EUA.

Ao Marcelo, por estar ao meu lado e me ajudar em todos os momentos, mesmo estando tão longe. Pelo mapa, por arrumar os gráficos e figuras e por tantas outras coisas que não conseguiria listar aqui. Por nunca me deixar desistir.

À minha família, por permanecer sempre ao meu lado torcendo pelo meu sucesso, especialmente à minha mãe, por me ajudar no que fosse preciso para que eu conseguisse realizar os meus sonhos.

A todos os indivíduos que doaram amostras de sangue para nossas pesquisas e tornaram possível não só o meu trabalho, mas o trabalho de todos no laboratório.

Às agências financiadoras, CAPES, CNPQ e Fundação Araucária.

Dedico este trabalho a todos os cientistas e futuros cientistas que trabalham com amor, dedicação e honestidade.

"A única abordagem científica séria é a que respeita a realidade: se esta for complexa, apresentá-la de forma simples só pode ser uma traição."

Albert Jacquard

RESUMO

CD80 e *CD86*, também denominados *B7.1* e *B7.2*, são genes proximamente ligados no cromossomo 3 que codificam glicoproteínas da superfamília das imunoglobulinas, expressas na superfície das células apresentadoras de antígeno. Essas moléculas participam na ativação e inibição das células T através da ligação aos receptores CD28 e CTLA-4. Nesse estudo foram analisados polimorfismos dos genes *CD80* e *CD86* com o objetivo de investigar a diversidade genética, microevolução e relevância funcional. Foram genotipados 1.124 indivíduos, incluindo brasileiros de ancestralidade predominantemente européia, mista africana e européia e japonesa, 5 populações ameríndias e africanos. As regiões promotoras de *CD80* e *CD86* foram sequenciadas e utilizadas em ensaios de gene repórter com luciferase em células HEK293T. As proteínas foram quantificadas por citometria de fluxo em monócitos, estimulados com quatro ligantes de TLR, de indivíduos com diferentes genótipos. Sítios de ligação de fatores de transcrição foram inferidos *in silico*. Todos os alelos analisados foram encontrados em africanos, o que sugere sua origem na África antes das migrações humanas para fora do continente. Cinco novos alelos da região promotora de *CD80* foram identificados e confirmados por clonagem e sequenciamento, sendo o promotor 2 o mais provável alelo ancestral. Foi encontrado um efeito de seleção balanceadora em descendentes de japoneses, nos quais todos os alelos apresentam frequências elevadas. O nucleotídeo -79 é monomórfico em 4 populações ameríndias, nas quais a presença do alelo -79G é provavelmente resultado de fluxo gênico. O haplótipo 4 de *CD80* apresentou expressão significativamente menor do que os demais devido ao SNP *g.-7T>C* (rs16829980), *g.5C>A* (rs41271391) e/ou *287A>T* (rs56124423). O SNP *g.-7T>C* está localizado no sítio de ligação da proteína ativadora AP-1 e de acordo com inferências *in silico* interfere na ligação. No gene *CD86*, dois SNPs foram analisados em 2.4 kb incluindo o promotor e a 5'UTR. O SNP *CD86 -298T>C* não interfere na expressão gênica segundo os experimentos com a luciferase. A quantidade das proteínas em monócitos diferiu significativamente entre os haplótipos de *CD80* e os alelos *-819T>C* (rs11575855) de *CD86* e também entre idosos e jovens e entre os quatro diferentes estímulos utilizados. Concluímos que polimorfismos na região 3' do promotor de *CD80* alteram significativamente a atividade do promotor. O SNP *CD86 -819T>C* altera os níveis proteicos de *CD86* na membrana de monócitos estimulados. As diferenças encontradas são provavelmente devidas a alterações na capacidade de ligação de fatores de transcrição.

Palavras-chave: CD80, CD86, polimorfismos, expressão, população

ABSTRACT

CD80 and *CD86*, also called *B7.1* and *B7.2*, are closely linked genes on chromosome 3 that code for glycoproteins of the immunoglobulin superfamily, expressed on the surface of antigen-presenting cells. These costimulatory molecules play essential roles for stimulation and inhibition of T cells through binding to CD28 and CTLA-4 receptors. In this study, *CD80* and *CD86* polymorphisms were analyzed to investigate the genetic diversity, microevolution, and the functional relevance of *CD80* and *CD86* promoter polymorphisms. We genotyped 1,124 individuals, including Brazilians of predominantly European, mixed African and European, and Japanese ancestry, 5 Amerindian populations, and an African sample. The *CD80* and *CD86* promoter regions were sequenced and functional studies were done using luciferase reporter assays in HEK293T cells, flow cytometry measurements of protein on TLR stimulated monocytes (using four different stimuli) of individuals with different genotypes, and *in silico* inferences of transcription factor binding. All variants were observed in Africans, which suggests their origin in Africa before the human migrations out of that continent. Five new *CD80* promoter alleles were identified and confirmed by cloning and sequencing, and promoter 2 is most likely the ancestral allele. There is evidence of an effect of balancing selection in Japanese-brazilians, where all alleles present high frequencies. Nucleotide-79 is monomorphic in 4 Amerindian populations, where the presence of the -79G allele is probably the result of gene flow from non-Amerindians. *CD80* haplotype 4 showed significantly lower promoter activity, caused by SNPs *g.-7T>C* (rs16829980), *g.5C>A* (rs41271391) and/or *287A>T* (rs56124423). The SNP *g.-7T>C* is located in the previously reported binding site of activating protein 1 (AP-1) and is predicted to interfere with AP-1 binding. In *CD86*, two SNPs were analyzed in the 2.4 kb including the promoter and 5'UTR of the gene. Luciferase measurements showed that the *CD86 -298T>C* SNP did not interfere in gene expression. The protein expression on monocytes differed significantly among *CD80* haplotypes and *CD86 -819T>C* (rs11575855) alleles, and also between age groups and the different stimuli given to the cells. We conclude that polymorphisms within the 3' end of the *CD80* promoter significantly affect the activity of the promoter. Further, the *CD86 -819T>C* SNP has a significant effect on protein levels on the membrane of stimulated monocytes. These differences are most likely caused by differential binding of transcription factors.

Key words: *CD80*, *CD86*, polymorphisms, expression, population

LISTA DE FIGURAS

Figura 1 - Mapa físico mostrando em destaque a localização das populações analisadas neste estudo	15
Figura 2 – Papel das moléculas B7 (CD80 e CD86), na regulação das células T.....	16
Figura 3 – Moléculas coestimuladoras das famílias B7 e CD28.....	17
Figura 4 - Representação da organização interna dos genes <i>CD80</i> e <i>CD86</i> , mostrando exons e introns com seus tamanhos em pares de base.....	19
Figura 5 – Sequência amplificada da região promotora e exon 1 do gene <i>CD80</i>	65
Figura 6 – Sequência amplificada da região promotora e exon 1 do gene <i>CD86</i>	66
Figura 7 - Gel de agarose 1% após corrida eletroforética de produtos de PCR do gene <i>CD80</i> corado com brometo de etídeo e fotografado sob luz ultravioleta	69
Figura 8 – Diagrama esquemático do vetor pSMART-HCAmp com <i>high-copy replication origins</i> e resistência a ampicillina	70
Figura 9 – Diagrama esquemático do vetor pGL4.18 [luc2P/Neo]	71
Figura 10 – Gel de agarose 1% após corrida eletroforética de plasmídeos extraídos de <i>E.coli</i> , digeridos com as enzimas <i>KpnI</i> e <i>SacI</i> , corado com brometo de etídeo e fotografado sob luz ultravioleta.....	71
Figura 11 – Gel de agarose 1% após corrida eletroforética de plasmídeos pGL4.18 digerido com as enzimas <i>KpnI</i> e <i>SacI</i> , corado com brometo de etídeo e fotografado sob luz ultravioleta	71
Figura 12 – Gel de agarose 1% após corrida eletroforética de plasmídeos pGL4.18 digerido com as enzimas <i>KpnI</i> e <i>SacI</i> , corado com brometo de etídeo e fotografado sob luz ultravioleta	72
Figura 13 – Sequência da região promotora do gene <i>CD80</i> mostrando a localização dos oligonucleotídeos iniciadores utilizados para as duas PCRs (setas em preto e setas em vermelhos), seguidas da amplificação do fragmento completo (setas das extremidades) realizadas para a produção do haplótipo 4 sem a inserção de 5 pb. Destado em cores a localização dos polimorfismos presentes na região.	73
Figura 14 – Gel de agarose 1% após corrida eletroforética dos haplótipos 2 e 4 do gene <i>CD80</i> digeridos para a produção dos haplótipos recombinantes 2/4 e 4/2, corado com brometo de etídeo e fotografado sob luz ultravioleta	74

SUMÁRIO

1	INTRODUÇÃO.....	11
	1.1 Populações	12
	1.2 Modelo de ativação das células T	16
	1.3 os genes <i>cd80</i> (<i>B7.1</i>) e <i>cd86</i> (<i>B7.2</i>)	17
2	OBJETIVOS E JUSTIFICATIVA	23
3	CAPÍTULO I.....	24
4	CAPÍTULO II.....	32
5	DISCUSSÃO	56
6	CONCLUSÕES	59
	REFERÊNCIAS	60
	APÊNDICE	64
	ANEXOS.....	75

1 INTRODUÇÃO

O estudo da diversidade de genes em populações possui uma série de aplicações, auxiliando no entendimento da evolução humana, da história demográfica das populações e da ação de fatores evolutivos, que pode ser observada através de sinais nas sequências genômicas. A mutação, o fluxo gênico, a deriva genética e a seleção podem agir de forma diferente e com diferentes intensidades em cada população. Assim, as frequências alélicas podem diferir entre populações, e alelos muito comuns em um grupo podem ser raros ou mesmo ausentes em outro. Os padrões de desequilíbrio de ligação também podem diferir entre populações, como um reflexo de suas histórias evolutivas.

A população brasileira é de origem bastante heterogênea, com contribuição de ameríndios, africanos e europeus da península ibérica, e posteriormente, de migrantes da Ásia e de outras regiões da Europa. O estudo de polimorfismos genéticos pode auxiliar na compreensão do surgimento e evolução do gene investigado, revelando evidências da ação de fatores evolutivos em cada uma das populações.

As populações indígenas são em geral populações isoladas, nas quais o fluxo gênico recente é bem menos intenso se comparado a populações urbanas. Os efeitos de gargalo de garrafa sofridos pelas populações indígenas durante sua história e a grande redução populacional recente durante a colonização europeia do continente americano resultaram em perda de diversidade genética nessas populações, observada em diversos locos do genoma (HEY, 2005).

As populações africanas, por outro lado, apresentam em geral grande diversidade e blocos menores de desequilíbrio de ligação. Isso ocorre porque nossa espécie surgiu no continente africano há cerca de 200 mil anos e começou a ocupar o restante do mundo há cerca de 100 mil anos, possivelmente a partir da região sudoeste africana, entre Namíbia e Angola. Portanto, as populações africanas são as mais antigas e mantiveram sua população efetiva relativamente grande durante todo esse período (TISHKOFF et al., 2009).

Os genes *CD80* e *CD86* são homólogos e estão ligados no braço longo do cromossomo 3. Codificam glicoproteínas do tipo I pertencentes à superfamília das imunoglobulinas, que compartilham os receptores CD28 e CTLA-4 e estão envolvidas na regulação da ativação das células T. A imunidade natural é a primeira linha de defesa contra microrganismos, porém as respostas específicas são fornecidas pela imunidade adquirida, que é composta por duas linhas de atuação: humoral e celular. A imunidade humoral é mediada por anticorpos, produzidos pelos linfócitos B. Já a imunidade celular é mediada pelos

linfócitos T (ou células T) e atua contra microrganismos intracelulares, destruindo-os ou destruindo as células infectadas, além de atuar também contra células próprias alteradas. Os linfócitos T reconhecem peptídeos antigênicos apenas quando ligados às moléculas de MHC do hospedeiro, que são expressas na superfície das células apresentadoras de antígeno (APC). Este reconhecimento é feito através do receptor de célula T (TCR). Para que ocorra a ativação dos linfócitos T, são necessários outros sinais, denominados coestimuladores, enviados pela mesma APC. O sinal mais bem conhecido ocorre pela ligação das moléculas B7 (CD80 e CD86), nas APCs, ao receptor CD28, nas células T.

Devido ao importante papel dos genes *CD80* e *CD86* no sistema imunológico, alguns polimorfismos e os níveis das proteínas CD80 e CD86 já foram investigados em estudos de associação a doenças autoimunes, alergias e resposta a transplantes (ABDALLAH et al., 2006; WONG et al., 2005a; WONG et al., 2005b; IP et al., 2006). Embora a região reguladora desses genes não esteja totalmente caracterizada, sabe-se que polimorfismos localizados na região promotora de genes podem interferir na regulação da expressão gênica, alterando os níveis de transcrição. Isso ocorre principalmente devido à alteração de sequências nas quais se ligam fatores de transcrição, proteínas nucleares que se ligam ao DNA ativando ou inibindo a expressão do gene. O objetivo do presente estudo foi a análise evolutiva de polimorfismos dos genes *CD80* e *CD86* e a investigação do impacto destes na expressão gênica.

1.1 POPULAÇÕES

A genética de populações estuda a origem e dinâmica da variação genética nas populações, e tem como um de seus objetivos compreender a ação dos fatores que determinam a evolução. Os polimorfismos genéticos são bastante informativos para este fim. Polimorfismo genético é a existência em um locus de pelo menos dois alelos em uma população, sendo que nenhum deles ocorre com frequência superior a 99%. Há vários tipos de polimorfismos, porém os mais comuns são os SNPs (*single nucleotide polymorphism* ou polimorfismo de nucleotídeo único), que consistem em troca, perda ou adição de um único nucleotídeo.

Embora a maior parte da variação genética seja encontrada dentro das populações (85%), as frequências alélicas variam entre as populações (8% entre populações locais e 7% entre os continentes). Estas variações ocorrem devido à ação de fatores evolutivos, de acordo com a história de cada população. Para compreender a história e evolução das populações é necessário estudar um grande número de polimorfismos. Para compreender a história dos

polimorfismos, é necessário estudar um grande número de populações (CAVALLI-SFORZA, 1998; CAVALLI-SFORZA e FELDMAN, 2003).

A população brasileira é miscigenada, composta por indivíduos de ascendências européia, africana sub-saariana, ameríndia e oriental, em diferentes proporções dependendo da região estudada. Na região sul, onde foi realizado o presente estudo, há predominância do componente europeu. A região metropolitana de Curitiba é composta por 84,65% de brancos, 11,34% de pardos, 2,49% de negros e 1,07% de amarelos (para o percentual restante não foi feita declaração). Estes dados mostram a origem bastante heterogênea de nossa população, porém esta classificação não revela os diferentes graus de mistura. Dois estudos analisaram o polimorfismo dos genes *HLA* na população de Curitiba. Um deles, realizado com genes *HLA* de classe I (*HLA-A*, *-B* e *C*), mostrou que não só para os brancos, mas também para os mulatos, o componente europeu é predominante. A contribuição de europeus, africanos e ameríndios foi, respectivamente, 94%, 3% e 3% para brancos e 57%, 39% e 4% para mulatos (BRAUN-PRADO et al., 2000). Outro estudo analisou genes *HLA* de classe I (*HLA-A*, *-B*, *C*) e II (*HLA-DR* e *-DQ*) e mostrou que o componente europeu é predominante entre os brancos (80,5%), sendo que a contribuição de africanos é de 12,5% e a de ameríndios é de 7,0%. Já para os mulatos o componente predominante foi o africano (49,5%), sendo as contribuições européia de 41,8% e ameríndia de 8,7% (PROBST et al., 2000).

Estudos genéticos de populações ameríndias, aliados a abordagens arqueológicas e linguísticas levaram à elaboração de hipóteses sobre o povoamento das Américas. A hipótese mais aceita hoje é a de que um pequeno número de indivíduos migraram da Sibéria para a América, durante a última glaciação, através do estreito de Bering (HEY, 2005). As migrações subsequentes foram da América do Norte em direção à América Central e do Sul. Durante as migrações, as populações passaram por deriva genética com sucessivos gargalos populacionais e efeitos de fundador. Análises de DNA mitocondrial, cromossomo Y e outras regiões polimórficas e mais recentemente análises genômicas utilizando microssatélites e SNPs espalhados por todo o genoma revelaram informações importantes sobre a provável data da primeira migração, número de migrações e rotas utilizadas (REICH et al., submetido; WANG et al.; 2007; SALZANO, 2002).

Através do estudo de genes *HLA*, de outros genes do MHC e de marcadores da região em populações indígenas, observou-se que a diversidade dos locos analisados em populações indígenas é menor do que a de outras populações, e o desequilíbrio de ligação entre os locos é maior (FAUCZ, 2000). Dentre as populações ameríndias foram encontrados alelos *HLA*

novos e exclusivos destas populações, que surgiram a partir dos alelos fundadores principalmente por eventos de conversão gênica (PARHAM et al., 1997).

Estima-se que mais de 1.000 povos indígenas (entre 2 e 4 milhões de pessoas) habitavam o Brasil na época da chegada dos europeus, mas hoje restaram apenas 230, totalizando 817.963 pessoas das quais 315.180 vivem em cidades e 502.783 em áreas rurais (0,42% da população brasileira segundo o Censo IBGE 2010), falantes de mais de 180 línguas e dialetos, distribuídos por 669 terras indígenas. Dentre esses, 51 mil são Guarani e 33 mil são Kaingang, as duas principais tribos indígenas do sul do Brasil, que possuem língua e cultura distintas, e mesmo quando habitam a mesma área indígena, moram em aldeias diferentes e tendem a permanecer endogâmicos (Funasa 2009 e Instituto Socioambiental).

Os Guarani eram nômades e viviam principalmente da agricultura. Habitavam uma grande área na região litorânea dos estados brasileiros do Rio Grande do Sul, Santa Catarina, Paraná, São Paulo e também no Mato Grosso do Sul, além de países vizinhos. Hoje vivem na região sul e nos estados de Mato Grosso do Sul, São Paulo, Rio de Janeiro, Espírito Santo, Pará e Tocantins e também na Argentina e no Paraguai, e são a maior etnia indígena brasileira. Sua língua é o Guarani, família Tupi-Guarani, tronco linguístico Tupi. As línguas dos Guarani analisados neste estudo são o M'byá (residentes na área indígena Rio das Cobras), Kaiowá (residentes em Amambai e Limão Verde, Mato Grosso do Sul) e Ñandeva (residentes em Amambai e Porto Lindo, MS) (Figura 1). A terra indígena localizada em Amambai, com 2.429 ha abriga aproximadamente 7106 pessoas entre Guarani Kaiowá e Ñandeva. Em Limão Verde vivem aproximadamente 460 indivíduos, e em Porto Lindo cerca de 1600 pessoas (Funasa, 2008).

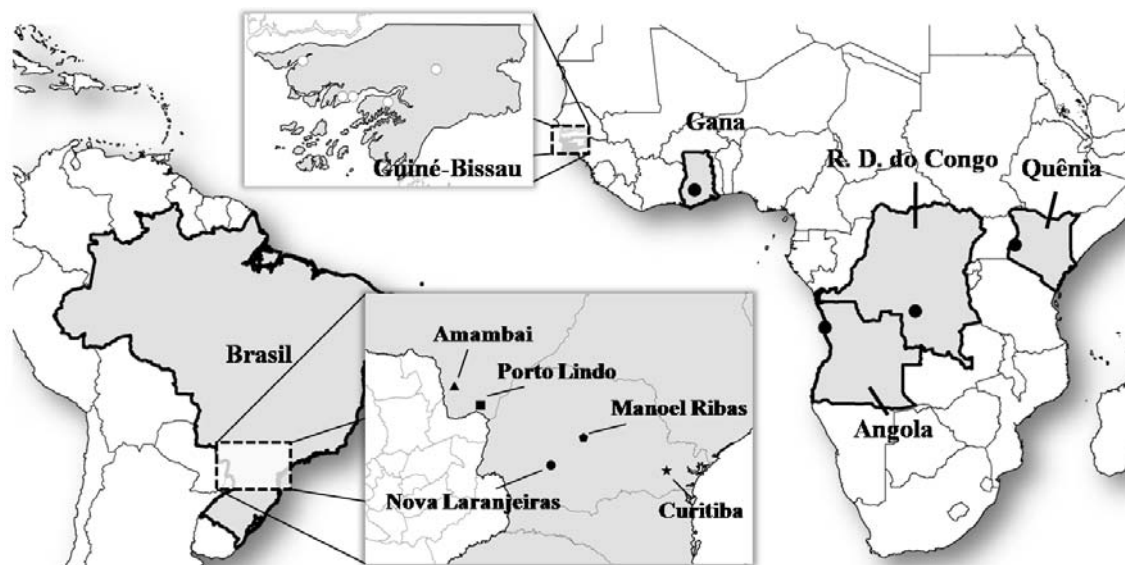


Figura 1 - Mapa físico mostrando em destaque a localização das populações analisadas neste estudo

Os Kaingang ocupavam grandes áreas no sul do Brasil (sempre na região central dos estados) e na Argentina. Viviam principalmente da agricultura, além da caça, pesca e coleta. Eram semi-nômades e guerreiros. Hoje são a terceira etnia indígena em tamanho da população e estão restritos a 32 terras indígenas, na região sul e no estado de São Paulo, muitas vezes compartilhadas com outras tribos. A tribo é dividida em dois grupos, Kame e Kairu, sendo que os indivíduos de um grupo devem se casar com os do grupo oposto, e os descendentes recebem a filiação paterna (as famílias são patrilineares). Sua língua é o Kaingang, classificado na família Jê, tronco linguístico Macro-Jê. Neste estudo foram analisados Kaingang residentes nas terras indígenas Rio das Cobras (a maior do Paraná em extensão territorial e em população, com 18.682 ha, localizada na cidade de Nova Laranjeiras), que abriga 2828 indivíduos entre Kaingang, Guarani e Xetá; e Ivaí (localizada na cidade de Manoel Ribas, Paraná, com 7.306 ha), que abriga 1510 Kaingang (Figura 1; FUNAI Chapecó, 2010).

Estudos realizados sobre as populações Kaingang e Guarani mostraram que, além da diferença entre populações indígenas e urbanas, elas são geneticamente diferentes entre si, apesar de atualmente conviverem em muitas áreas indígenas (PETZL-ERLER et al., 1993; WEG-REMERS et al., 1997). A taxa de miscigenação com populações não indígenas foi estimada em 4% para Guarani M'byá, 14% para Guarani Ñandeva, ausente para Guarani

Kaiowá e 7% para Kaingang através da análise de genes HLA de classe II (*HLA-DRB1*, *-DQA1* and *-DQB1*) (TSUNETO et al., 2003).

1.2 MODELO DE ATIVAÇÃO DAS CÉLULAS T

Para que ocorra a ativação, proliferação e diferenciação dos linfócitos T virgens em células efectoras, ao menos dois sinais enviados pelas células apresentadoras de antígeno profissionais são de fundamental importância (Figura 2). O primeiro sinal é antígeno-específico e é dado pelo reconhecimento do complexo peptídeo/moléculas MHC de classe II (Complexo Principal de Histocompatibilidade) nas APC pelo receptor de célula T (TCR) e pelo coreceptor CD4. O segundo sinal deriva da interação entre as moléculas coestimuladoras B7 (sendo as mais bem caracterizadas CD80 e CD86) nas APC e o receptor CD28 nos linfócitos T. Sem o segundo sinal não ocorre ativação e as células T tornam-se anérgicas. Após a ativação, a célula T passa a expressar uma variedade de moléculas que contribuem para a sustentação ou modificação dos sinais coestimuladores.

O sinal regulador negativo, que controla a resposta proliferativa e promove a tolerância periférica, prevenindo a autoimunidade, é dado pela ligação das moléculas B7 ao receptor CTLA-4 (*citotoxic T-lymphocyte antigen-4*), expresso na membrana dos linfócitos T ativos (Figura 2). Esta ligação ocorre com cerca de 20 vezes maior afinidade do que a ligação de B7 a CD28 e é essencial para a limitação da resposta proliferativa das células T. A regulação negativa pode ser promovida não apenas por sinais negativos enviados através da ligação de CTLA-4, mas também pela exclusão dos complexos CD28/B7 e/ou por competição direta pela ligação das moléculas B7 (BHATIA et al., 2006). Assim, as moléculas CD28 e CTLA-4, através da ligação às moléculas B7, regulam as respostas imunes contra antígenos próprios e estranhos controlando a ativação antígeno-específica das células T (REEVES et al., 1997).

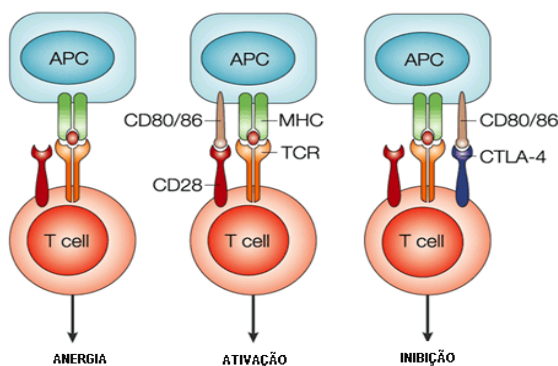


Figura 2 – Papel das moléculas B7 (CD80 e CD86), na regulação das células T

Fonte: Alegre, Frauwirth e Thompson, 2001

Embora as vias de ativação das células T que envolvem as moléculas coestimuladoras CD80 e CD86 sejam as mais conhecidas, há outras moléculas que atuam neste processo (Figura 3). Enquanto CD80 e CD86 possuem expressão restrita a células linfóides, ICOS-L, PD-L1, B7-H3 e B7-H4 são expressas também em células não linfóides, o que sugere que estas moléculas têm papéis diferentes no sistema imune (COLLINS, LING e CARRENO, 2005).

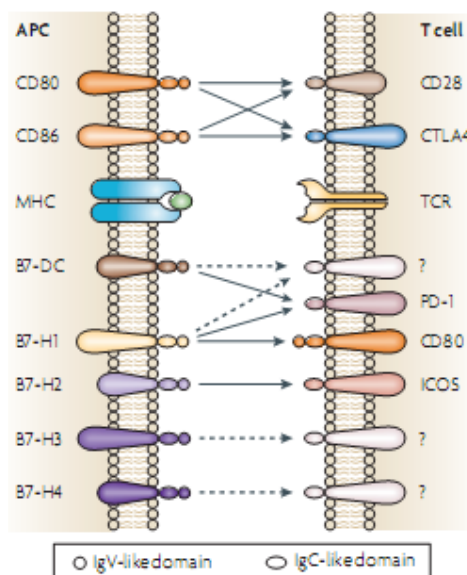


Figura 3 – Moléculas coestimuladoras das famílias B7 e CD28

Fonte: ZOU e CHEN, 2008.

1.3 OS GENES *CD80* (*B7.1*) E *CD86* (*B7.2*)

Os genes *CD80* e *CD86* são homólogos e estão ligados no braço longo do cromossomo 3 (3q21). Codificam as glicoproteínas tipo 1, estruturalmente relacionadas, CD80 e CD86, também denominadas B7.1 e B7.2, membros da superfamília das imunoglobulinas. Essas moléculas são expressas na membrana de células apresentadoras de antígenos como células dendríticas, monócitos e células B ativadas e possuem um papel essencial na modulação da ativação dos linfócitos (ou células) T (CHAMBERS, 2001).

As moléculas CD80 e CD86 são ambas capazes de promover a coestimulação e a proliferação das células T, mas elas não são idênticas: o padrão de expressão difere (ao contrário de CD80, CD86 tem expressão constitutiva), assim como as características bioquímicas e os estados oligoméricos. CD80 forma dímeros e CD86 forma monômeros na

superfície das APC. CD86 liga CD28 de forma mais eficaz do que CD80 e é considerado o principal ligante de CD28; que também apresenta expressão constitutiva, enquanto CTLA-4 tem maior afinidade por CD80 do que por CD86, sendo ambos de expressão induzida após a ativação das células T (COLLINS et al., 2002). Essas moléculas também modulam as respostas Th1 e Th2. Enquanto CD80 favorece a produção de citocinas Th1, CD86 favorece o tipo Th2. Além disso, CD80 e CD86 apresentam domínios intracelulares distintos e podem mediar diferentes vias de transdução de sinal nas APC (BHATIA et al., 2006; SAMSON, MANZOTTI e ZHENG, 2003; SUVAS et al., 2002).

O gene *CD80* contém sete exons e se estende por cerca de 32 kb de DNA (Figura 4). Codifica uma glicoproteína de mesmo nome, que apresenta dois domínios extracelulares estruturalmente similares aos das imunoglobulinas, codificados pelos exons 3 e 4, uma região hidrofóbica transmembrana codificada pelo exon 5 e um curto domínio citoplasmático correspondente ao exon 6. Parte do exon 6 e o exon 7 correspondem à região 5'UTR. O exon 1 não é traduzido e o exon 2 contém o códon de iniciação AUG e codifica o peptídeo sinal (SELVAKUMAR et al., 1992; BHATIA et al., 2006; SAMSON, MANZOTTI e ZHENG, 2003).

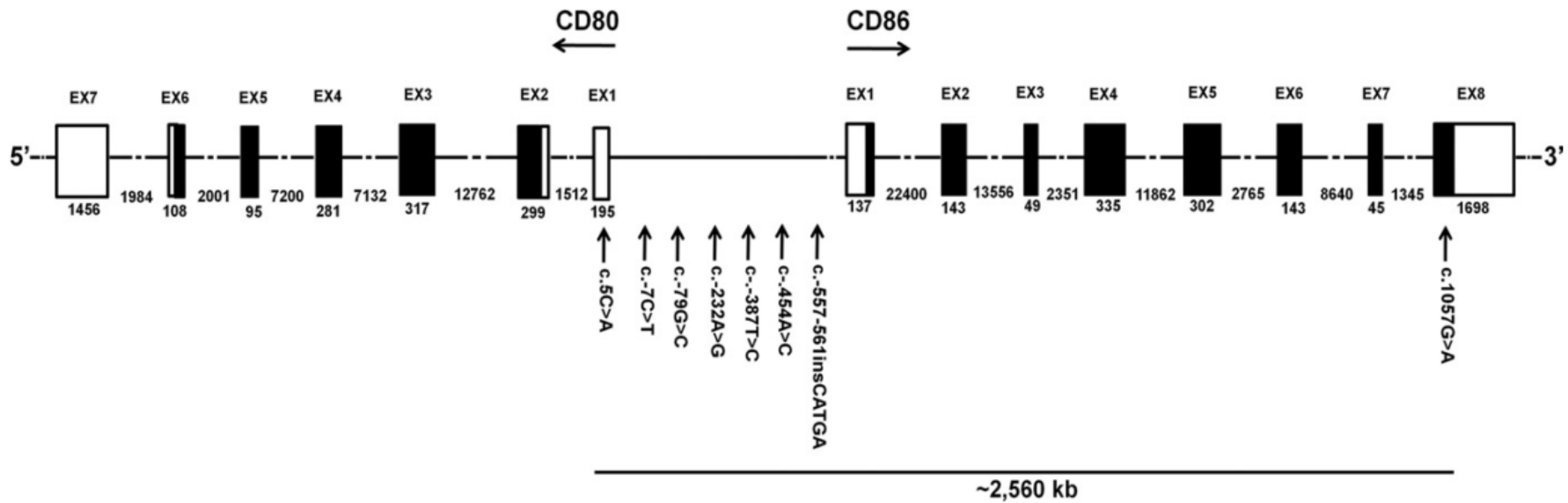


Figura 4 - Representação da organização interna dos genes *CD80* e *CD86*, mostrando exons (blocos) e introns (linhas) com seus tamanhos em pares de base e a localização dos SNPs investigados. Blocos em branco indicam regiões não traduzidas e em preto indicam região codificadora. O pontilhado indica que a sequência foi reduzida, a figura não está representada em escala.

Fonte: DALLA-COSTA et al., 2010.

Seis polimorfismos na região promotora deste gene, sendo cinco SNPs (polimorfismos de nucleotídeo único), *c.-454C>A* (rs68180496), *c.-387T>C* (rs1880661), *c.-232G>A* (rs41271393), *c.-79G>C* (rs16829984), *c.-7C>T* (rs16829980) e um indel (polimorfismo de inserção/deleção) de 5pb, *c.-557_-561insCATGA* e um SNP na região 5' não traduzida, posição *c.5C>A* (rs41271391), já foram analisados na população japonesa, na qual foram descritos (MATSUSHITA et al., 2000) e também em australianos de origem europeia (TEUTSCH et al., 2004), com exceção das posições -558 e +5, analisadas apenas em japoneses. Apenas quatro haplótipos foram encontrados para a combinação dos sete polimorfismos localizados na região reguladora do gene *CD80* em japoneses e australianos e foi sugerido que esses haplótipos estão sob efeito de seleção balanceadora (MATSUSHITA et al., 2000; TEUTSCH et al., 2004).

A regulação da expressão dos genes *CD80* e *CD86* não é bem caracterizada, porém algumas regiões de regulação já foram identificadas. Zhao et al. (1996) encontraram evidências de uma região acentuadora localizada aproximadamente 3 kb a montante do primeiro exon do gene *CD80*, que determina a expressão tecido-específica e inducibilidade por vários estímulos, incluindo um sítio de ligação para NF- κ B, necessário para a ligação da proteína p50. Mutações nesse elemento resultaram em mais de 70% de redução da atividade acentuadora. Usando o programa GCG no banco de dados GenBank, eles encontraram duas sequências repetitivas Alu como elementos reguladores negativos, e também dois elementos TATAA, nas posições -2608 e -2501 relativas ao sítio de início de transcrição. Fong, Wu e Kipps (1996) identificaram um elemento regulador positivo na região promotora de *CD80* chamado B7-RE, localizado entre -60 e -47 e também um sítio de ligação da proteína AP-1 na região onde posteriormente Matsushita et al. (2000) identificaram o SNP *-7T>C*, que pode interromper a ligação de acordo com o programa TFSEARCH. Lim et al. (2005) identificaram um sítio de ligação para o fator de regulação IFN 7 (IRF-7) localizado entre -84 e -72 pb a montante do sítio de início da transcrição e mostraram que mutações nessa sequência impedem a transcrição.

O gene *CD86* compreende um segmento de mais de 22 kb de DNA e contém oito exons (Figura 2). Há um sítio de recomposição alternativa entre os exons 1 e 2. O exon 3 corresponde ao peptídeo sinal e os exons 4 e 5 aos domínios extracelulares. O exon 6 codifica a região transmembrana e, junto com os exons 7 e 8, a cauda citoplasmática. (BHATIA et al., 2006; SAMSON, MANZOTTI e ZHENG, 2003). A molécula CD86 possui três sítios de fosforilação pela proteína quinase C (PKC) em sua cauda citoplasmática e possivelmente apresenta função de transdução de sinal nas células B (JUNE et al., 1994).

Uma forma solúvel de CD86 (CD86deltaTM) é produzida por recomposição alternativa que resulta na remoção do exon 6 e conseqüentemente na ausência da região transmembrana. Esta molécula foi encontrada no soro humano e é funcional, sendo capaz de transmitir sinal coestimulador às células T de memória (JEANNIN et al., 2000). Outra molécula de CD86 (CD86ΔEC), na qual falta o sítio de ligação aos receptores CD28 e CTLA-4, é produzida por recomposição alternativa com a ausência dos exons 4 e 5 no RNAm, que codificam os domínios extracelulares semelhantes a imonoglobulina, nos quais localizam-se os sítios de ligação aos receptores CD28 e CTLA-4. Formas solúveis de CD80 também já foram descritas no soro, plasma ou líquido sinovial de pacientes de diversas doenças, porém não é conhecido se são formas produzidas por recomposição alternativa do RNA mensageiro ou por clivagem enzimática de moléculas completas (KAKOULIDOU et al., 2007). Essas moléculas podem interferir na ligação de CD80 e CD86 a CD28, agindo como reguladoras negativas da ativação das células T. As três variações da molécula CD86 são produzidas por monócitos não ativados, porém após a ativação, a molécula completa é preferencialmente expressa (MAGISTRELLI et al., 2001). Além disso, CD86 tem sítios de fosforilação pela proteína quinase C no domínio citoplasmático e é possível que apresente função de transdução de sinal nas células B (JUNE et al., 1994).

Cinco polimorfismos foram bem caracterizados na região reguladora do gene *CD86*, três no promotor e dois na região 5'UTR. Turpeinen et al. (2002) descreveram o SNP -819T>C (rs11575855) na região promotora, 51A>G no exon 1 (rs11575853) e 19G>A no exon 2 (rs9282641). A deleção -359delAAG, localizada na região 5'UTR, a montante do início do exon 2, de acordo com o programa TRANSFAC (WINGENDER et al., 2000) não altera sítios de ligação para fatores de transcrição (TEUTSCH et al., 2004). O alelo G do SNP -3479G>T (rs2715267) está associado com esclerose sistêmica e analisado por ensaios de retardamento de migração de banda em eletroforese (EMSA) tem menor afinidade na ligação por proteínas do núcleo. O alelo T possivelmente cria sítios GATA e TATA box (TBP) do nucleotídeo -3469 ao -3486 de acordo com o programa TRANSFAC (ABDALLAH et al., 2006). Possíveis sequências de ligação para fatores de transcrição já foram identificadas no gene *CD86* com a utilização do programa Transfast (Gesellschaft für Biotechnologische Forschung, Braunschweig, Germany), incluindo pelo menos dois sítios de ligação para NF-κB localizados a -612 (5' GGGACATCTA 3') e -238 (5' GGGAATTTC 3') (LI et al., 1999).

1.4 AVALIAÇÃO DO IMPACTO FUNCIONAL DE POLIMORFISMOS

Polimorfismos localizados em regiões promotoras podem potencialmente interferir na regulação da expressão gênica. Além de poderem alterar o padrão de metilação do DNA, o efeito mais direto é a alteração de sítios de ligação para fatores de transcrição. Mais comumente os polimorfismos alteram a afinidade da ligação DNA-proteína, mas em alguns casos podem chegar a impedir a ligação.

Fatores de transcrição são proteínas que se ligam ao DNA através do reconhecimento de sequências específicas e podem aumentar ou reduzir a expressão do gene. As sequências de ligação para fatores de transcrição são geralmente curtas (de 4 a 20 pb) e degeneradas, sendo que o grau de conservação de cada base varia.

A conservação de sequências entre espécies diferentes é considerada um indicador importante da funcionalidade, embora sequências que apresentam funções importantes não sejam necessariamente conservadas (a função pode ser importante em apenas uma espécie e não nas demais, por exemplo). Com o alinhamento de múltiplas sequências ortólogas é possível detectar curtos sítios de ligação para fatores de transcrição que são significativamente mais conservados do que seria esperado ao acaso. A comparação entre sequências genômicas ortólogas para identificar regiões conservadas entre espécies que têm provavelmente importância funcional é chamada de “pegada filogenética”, do inglês *phylogenetic footprinting* (FICKETT e WASSERMAN, 2000).

A sequência de ligação de um fator de transcrição permite alguma variação, criando um espectro de afinidades de ligação, que é representado em um gráfico onde cada nucleotídeo da sequência aparece em tamanho proporcional à sua importância para a ligação à proteína. Muitas sequências reguladoras já foram descritas e depositadas em bancos de dados na internet, de acesso livre. Como são conhecidas as sequências e as proteínas que se ligam a elas, é possível prever a partir de uma sequência em particular, quais fatores de transcrição poderiam regular a expressão do gene investigado. Comparando o resultado entre sequências de alelos diferentes é possível prever o impacto do polimorfismo na ligação à proteína. Porém é necessário lembrar que alguns desses fatores de transcrição podem não estar presentes no tecido ou tipo celular específico no qual o gene é expresso.

Além das análises por bioinformática, existem técnicas que permitem a realização de experimentos testando o nível de expressão do gene. Uma delas utiliza a clonagem em plasmídeo contendo um gene repórter, no qual é inserido o promotor do gene de interesse. Assim é possível comparar o nível de expressão resultante dos diferentes alelos.

2 OBJETIVOS E JUSTIFICATIVA

CD80 e CD86 possuem funções essenciais na regulação da ativação das células T como moléculas coestimuladoras e coinibidoras. Polimorfismos nos genes *CD80* e *CD86* e níveis das proteínas no soro humano já foram relacionados a doenças autoimunes, a alergias e à resposta a transplantes. Embora eles sejam genes ligados homólogos e altamente similares, eles podem estar sob diferentes regimes de seleção natural, como evidenciado pela observação de que essas moléculas apresentam importantes funções que são apenas parcialmente redundantes nas respostas imunes. Entretanto, pouco é conhecido sobre a regulação da expressão dos genes *CD80* e *CD86*, especialmente sobre o impacto dos polimorfismos localizados em regiões reguladoras. Também a história evolutiva destes genes e de seus polimorfismos ainda é pouco conhecida.

Polimorfismos localizados na região promotora dos genes *CD80* e *CD86* podem alterar a afinidade de ligação de proteínas reguladoras do núcleo, como fatores de transcrição ativadores ou inibidores. Portanto, alelos podem diferir na afinidade por proteínas reguladoras, resultando em variação no nível de expressão, com efeitos funcionais. A hipótese desse trabalho é que existam polimorfismos e haplótipos ainda não descritos nos genes *CD80* e *CD86*, uma vez que foram estudados apenas em europeus e japoneses até o presente, e que algumas das variações interfiram na regulação da expressão gênica, alterando a afinidade de ligação de fatores de transcrição.

O objetivo deste estudo foi descrever a diversidade e verificar a relevância funcional das variações na região promotora dos genes *CD80* e *CD86* para a transcrição e analisar a origem e história evolutiva de polimorfismos dos genes *CD80* e *CD86*, buscando evidências da ação de fatores evolutivos nestas sequências genômicas.

Os objetivos específicos foram:

- verificar a existência de polimorfismos ainda não descritos na região promotora dos genes *CD80* e *CD86*;
- verificar se os polimorfismos da região promotora dos genes *CD80* e *CD86* apresentam evidências da ação de fatores evolutivos;
- verificar se os diferentes alelos da região promotora dos genes *CD80* e *CD86* resultam em diferentes níveis de expressão e identificar quais fatores de transcrição podem ser responsáveis por essas diferenças.

3 CAPÍTULO I



CD80 and CD86 polymorphisms in populations of various ancestries: 5 new CD80 promoter alleles

Márcia Holsbach Beltrame^a, Márcia Regina Pincerati^{a,†}, Ricardo Dalla-Costa^{a,‡}, Roseli Wassem^a, Karen Francine Köhler^b, Eleidi Alice Chautard-Freire-Maia^a, Luiza Tamie Tsuneto^c, Maria Luiza Petzl-Erler^{a,*}

^a Human Molecular Genetics Laboratory, Department of Genetics, Federal University of Paraná, 81531-980 Curitiba, Brazil

^b Laboratory of Immunology, Heart Institute (InCor), University of São Paulo, 05403-001 São Paulo, Brazil

^c Department of Basic Health Sciences, State University of Maringá, 87020-900 Maringá, Brazil

ARTICLE INFO

Article history:

Received 12 July 2011

Accepted 7 October 2011

Available online 22 October 2011

Keywords:

B7

Single nucleotide polymorphisms (SNP)

Haplotypes

Linkage disequilibrium

Amerindians

ABSTRACT

CD80 and CD86 are closely linked genes on chromosome 3 that code for glycoproteins of the immunoglobulin superfamily, expressed on the surface of antigen-presenting cells. These costimulatory molecules play essential roles for stimulation and inhibition of T cells through binding to CD28 and CTLA-4 receptors. In this study, CD80 promoter and CD86 exon 8 polymorphisms were analyzed to investigate the genetic diversity and microevolution of the 2 genes. We genotyped 1,124 individuals, including Brazilians of predominantly European, mixed African and European, and Japanese ancestry, 5 Amerindian populations, and an African sample. All variants were observed in Africans, which suggests their origin in Africa before the human migrations out of that continent. Five new CD80 promoter alleles were identified and confirmed by cloning and sequencing, and promoter 2 is most likely the ancestral allele. Nucleotide -79 is monomorphic in 4 Amerindian populations, where the presence of the -79 G allele is probably the result of gene flow from non-Amerindians.

© 2012 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.

1. Introduction

CD80 and CD86, also called B7.1 and B7.2, are closely linked genes on chromosome 3 (3q21) that code for structurally related type 1 glycoproteins, members of the immunoglobulin superfamily. CD80 and CD86 are transmembrane molecules expressed on the surface of antigen-presenting cells (APC), and play essential roles in T-cell modulation. The first signal required for the activation of T cells is antigen specific, based on recognition of the human leukocyte antigen (HLA)/peptide complex on APC via the T-cell receptor. The best characterized costimulatory signal derives from the interaction between B7 molecules CD80 and CD86, on APC, and the CD28 receptor on the T cell. The interaction of B7 with the CTLA-4 receptor gives the negative regulatory signal, which promotes peripheral T-cell tolerance [1,2].

Although both CD80 and CD86 are able to costimulate T-cell proliferation, they exhibit different patterns of expression, biochemical features, and oligomeric states. CD86 is thought to be the major CD28 ligand, whereas CTLA-4 has a higher affinity for CD80 than for CD86 [3]. In addition, whereas CD80 favors a T helper

response, CD86 favors T helper 2 type cytokine production. They also have distinct intracellular domains and may mediate differential signal transduction in APC [4–6].

Allelic frequencies of genetic variants may differ within human populations. Patterns of linkage disequilibrium (LD) may also vary among populations with different evolutionary histories, especially between modern Africans and populations distantly related to them [7]. Comparing the diversity observed within continents, Africa usually has the highest levels of polymorphism and the lowest values of LD. In addition, populations that went through bottleneck events, such as Amerindians [8], usually exhibit the highest values of LD and the lowest genetic diversity [9].

Although CD80 and CD86 play a crucial role in the immune system, the polymorphisms analyzed in this study have been poorly investigated. In this work Africans, Guarani and Kaingang Amerindians, and Brazilians of Japanese, European, and mixed African and European ancestry were analyzed. The frequencies of CD80 promoter and CD86 exon 8 variants and the LD between them were described to access information about the diversity and evolution of these polymorphisms.

2. Subjects and methods

2.1. Population samples

A total of 1,124 individuals from 4 urban and 5 Native American populations were genotyped for CD80 and CD86 polymorphisms.

* Corresponding author.

E-mail address: perler@ufpr.br (M. L. Petzl-Erler).

† Current address: Department of Genetics and Evolutionary Biology, University of São Paulo, Rua do Matão 277, Cidade Universitária, 05508-090 São Paulo, Brazil.

‡ Current address: Life Technologies, Avenida do Café 277, 1º Andar, Torre A, Bairro Vila Guarani, 04311-000 São Paulo, Brazil.

The urban Brazilian population sample, composed of genetically unrelated healthy individuals from the general population, was collected in Curitiba, in the state of Paraná (25°25' S, 49°16' W). The population is admixed but still retains some structure regarding the ancestry of individuals. Therefore, the sample was subdivided according to predominant ancestry into European (EU-B, $n = 158$), mixed African and European (AF-B, $n = 274$), and Japanese (JA-B, $n = 101$). In 2 previous papers as well as unpublished data from our group, the admixture of our population was estimated using HLA class I and II as ancestry markers. EU-B and AF-B had 87 and 50% of European, 8 and 44% of sub-Saharan African, and 5 and 6% of Amerindian ancestry, respectively [10,11]. JA-B reported exclusively Japanese ancestry. Also, a sample of 16 genetically unrelated African individuals was collected in Curitiba: 7 from Guinea-Bissau, 7 from Angola, 1 from Ghana, and 1 from Cameroon.

The Amerindian populations were Guarani Kaiowá (GKW, $n = 161$) from Limão Verde (23°12' S, 55°06' W) and Amambai (23°06' S, 55°12' W), Guarani Nandeva (GND, $n = 84$) from Amambai and Porto Lindo (23°48' S, 54°30' W), all from the state of Mato Grosso do Sul, and Guarani M'byá (GRC, $n = 89$) from Rio das Cobras (25°18' S, 52°32' W) in Nova Laranjeiras, in the state of Paraná. These 3 Guarani subgroups differ in various aspects of their culture. Kaingang were from Rio das Cobras (KRC, $n = 110$) and from Ivaí (24°30' S, 51°40' W) in Manoel Ribas, also in the state of Paraná (KIV, $n = 131$). Guarani populations are from the linguistic family Tupi-Guarani, which belongs to the Tupi branch, whereas Kaingang are from the linguistic family Macro-Jê, Jê branch. Although nowadays they live close to each other in many indigenous areas, they remain remarkably distinct in culture and genetic polymorphism. A common feature of the Kaingang and Guarani groups is the isolation from non-Amerindian populations. More detailed description of these populations can be found in previous publications [12,13].

The individuals gave their informed consent and the study was approved by the Committee for Ethics in Human Research from the Federal University of Paraná.

2.2. Genotyping

Genomic DNA was obtained from peripheral blood by salting-out [14] and phenol-chloroform-isoamyl alcohol [15] methods. Six single-nucleotide polymorphisms (SNPs), $g-454C>A$ (rs68180496), $g-387T>C$ (rs1880661), $g-232G>A$ (rs41271393), $g-79G>C$ (rs16829984), $g-7T>C$ (rs16829980), and $g.5C>A$ (rs41271391); 1 indel (insertion/deletion polymorphism), $g-557-561insCATGA$ (-558ins), in the promoter region of the *CD80* gene; and 1 SNP in exon 8 of the *CD86* gene, $g.1057G>A$ (rs1129055), were analyzed by polymerase chain reaction (PCR) followed by hybridization with sequence-specific oligonucleotide probes.

A fragment encompassing 690 bp of the *CD80* promoter region was amplified with the following PCR conditions: 80 ng of genomic DNA, 1× PCR buffer, 0.2 mM dNTP, 0.4 μM of each primer, and 0.75 U of Tth DNA polymerase (Biotools, Madrid, Spain) in 40 μL of reaction mixture. PCR started with 95°C for 5 minutes; followed by 9 cycles of 95°C for 30 seconds, 60°C for 45 seconds, 72°C for 45 seconds, and 26 cycles of 95°C for 30 seconds, 65°C for 45 seconds and 72°C for 45 seconds; with a final step at 72°C for 10 minutes. A segment of 166 bp of exon 8 of the *CD86* gene was amplified in a 10-μL reaction mixture set up to contain 20 ng of genomic DNA, 1× PCR buffer, 0.2 mM dNTP, 0.4 μM of each primer, 3 mM MgCl₂, and 0.2 U of Platinum Taq DNA polymerase (Invitrogen Corporation, Carlsbad, CA). PCR conditions were 95°C for 3 minutes; followed by 34 cycles of 95°C for 15 seconds, 56.7°C for 30 seconds, and 72°C for 40 seconds; with a final step at 72°C for 10 minutes.

PCR products were denatured and cross-linked on a nylon membrane. Dot blot hybridization was performed with biotin-labeled probes as previously described [16]. Probes used in genotyping (Table 1) were de-

Table 1
Oligonucleotide probes and high-stringency washing temperatures optimized for detection of the *CD80* promoter and *CD86* exon 8 polymorphisms by PCR with sequence-specific oligonucleotide probes

	Sequence (5'–3')	T (°C)
<i>CD80</i>		
	<i>c-557-561insCATGA</i>	<i>del</i> : GGTGGCAACCTAGTT <i>ins</i> : TGGCATGACAACCTTA
<i>c-454C>A</i>		
		C: TGTCACAGCTGGCTG A: TGTCACAGCTGGCTG
<i>c-387T>C</i>		
		T: GGTGCTCTGTGAGC C: GGTGCTCTGTGAGC
<i>c-232G>A</i>		
		G: ACACAGCAAGGCTAG A: ACACAGCAAGGCTAG
<i>c-79C>G</i>		
		G: AAAAGCAAGTGAAGAA C: AAAAGCAACTAGAAGAA
<i>c-7T>C</i>		
		T: TACTGAGTGAACCTCA C: TACTGAGTGAACCTCA
<i>c.5C>A</i>		
		C: CTCAAACCCCTCTGTA A: CTCAAACACTCTGTA
Control	TCACTCAGGGGCT	42
<i>CD86</i>		
	<i>c.1057G>A</i>	A: TGATGAAACCCAGCG G: TGATGAAAGCCAGCG
Control	GACATCTCATCGGA	48

signed based on the sequence of the human *CD80* and *CD86* genes (GenBank). Primers were described previously [17].

2.3. Cloning and sequencing

The new *CD80* promoter alleles were confirmed by cloning and sequencing the segment encompassing the promoter polymorphisms, from nucleotide -637 to 24 (chr3: 119278409–119279098). PCR products were cloned in *Escherichia coli* DH5-α competent cells using the PTZ57R/T vector (InsTAclone, Fermentas, Ontario, Canada) or TOPO (Invitrogen), and *E. coli* cells were grown in Luria broth-agar medium (Invitrogen Corporation, Carlsbad, CA) containing ampicillin (200 μg/mL). Plasmid vectors were extracted using the QIAprep Spin Miniprep kit (Qiagen Inc., Duesseldorf, Germany). The sequencing reaction was performed using M13 or Universal primers and chromatograms were visually analyzed in Bioedit [18] to confirm the DNA sequences.

2.4. Statistical analysis

Allelic and genotypic frequencies were estimated by direct counting. Hardy-Weinberg equilibrium (HW) was assessed using a modified version of the Markov-chain random walk algorithm [19]. Haplotypic frequencies were obtained using the ELB algorithm. Comparisons of the allele frequencies between the population samples were performed using contingency tables in RXC [20]. The p value of 0.05 was adopted as the significance limit for all statistical tests. These analyses, the estimates of the LD parameters D' , r^2 , and p , and the analysis of molecular variance were performed using the Arlequin 3.1 software package [21]. The relationship between haplotypes was assessed with the median-joining algorithm in Network 4.516 software. The dendrogram was constructed with the neighbor-joining method [22] using the modified Cavalli-Sforza chord measure for genetic distances [23] in the DISPAN software [24].

3. Results and discussion

3.1. Polymorphisms

Allelic and genotypic frequencies are presented in Table 2. *CD80* alleles -557-561**del*, -454°C, -232°G, -79°C, -7°T, and 5°C were the most common alleles in all populations. The *CD80* -387°C allele was the most frequent in EU-B and in Amerindians, whereas in AF-B and JA-B the -387°T allele was the most common.

Nucleotide -79 in the *CD80* promoter is polymorphic all over the "Old World," as indicated by the presence of this SNP in Africans,

Table 2
Allelic and genotypic frequencies (%) of the CD80 promoter variants and CD86 exon 8 single-nucleotide polymorphisms

	AF-B	EU-B	JA-B	GKW	GND	GRC	KRC	KIV
-557_-561ins	(n = 267)	(n = 154)	(n = 92)	(n = 161)	(n = 84)	(n = 88)	(n = 108)	(n = 129)
del	75.5	82.5	81.0	66.1	86.3	97.7	88.0	91.1
ins	24.5	17.5	19.0	33.9	13.7	2.3	12.0	8.9
del/del	57.3	69.5	66.3	45.3	73.8	95.5	76.9	82.9
del/ins	36.3	26.0	29.3	41.6	25.0	4.5	22.2	16.3
ins/ins	6.4	4.5	4.3	13.0	1.2	0.0	0.9	0.8
p-HW	0.74	0.26	0.73	0.38	1	1	1	1
-454C>A	(n = 259)	(n = 145)	(n = 90)	(n = 161)	(n = 84)	(n = 77)	(n = 102)	(n = 121)
C	74.9	81.7	80.6	66.1	86.3	97.4	87.3	90.9
A	25.1	18.3	19.4	33.9	13.7	2.6	12.7	9.1
C/C	56.0	68.3	65.6	45.3	73.8	94.8	75.5	82.6
C/A	37.8	26.9	30.0	41.6	25.0	5.2	23.5	16.5
A/A	6.2	4.8	4.4	13.0	1.2	0.0	1.0	0.8
p-HW	1	0.26	0.74	0.38	1	1	1	1
-387T>C	(n = 266)	(n = 158)	(n = 92)	(n = 161)	(n = 84)	(n = 86)	(n = 107)	(n = 127)
C	35.7	53.2	25.0	56.5	73.8	73.3	77.6	85.8
T	64.3	46.8	75.0	43.5	26.2	26.7	22.4	14.2
C/C	13.5	28.5	10.9	36.0	51.2	59.3	58.9	74.0
C/T	44.4	49.4	28.3	41.0	45.2	27.9	37.4	23.6
T/T	42.1	22.1	60.9	23.0	3.6	12.8	3.7	2.4
p-HW	0.60	1	0.02	0.04	0.16	0.01	0.58	0.71
-232G>A	(n = 263)	(n = 153)	(n = 92)	(n = 161)	(n = 84)	(n = 82)	(n = 106)	(n = 126)
G	75.9	82.4	81.0	66.1	86.3	97.6	88.2	90.9
A	24.1	17.6	19.0	33.9	13.7	2.4	11.8	9.1
G/G	57.0	69.3	66.3	45.3	73.8	95.1	77.4	82.5
G/A	37.6	26.1	29.3	41.6	25.0	4.9	21.7	16.7
A/A	5.3	4.6	4.3	13.0	1.2	0.0	0.9	0.8
p-HW	0.74	0.26	0.73	0.38	1	1	1	1
-79C>G	(n = 263)	(n = 152)	(n = 91)	(n = 161)	(n = 84)	(n = 82)	(n = 106)	(n = 124)
C	83.5	90.1	69.8	100	99.4	98.8	100	99.6
G	16.5	9.9	30.2	0.0	0.6	1.2	0.0	0.4
C/C	70.0	80.3	47.2	100	98.8	97.6	100	99.2
C/G	27.0	19.7	45.1	0.0	1.2	2.4	0.0	0.8
G/G	3.0	0.0	7.7	0.0	0.0	0.0	0.0	0.0
p-HW	0.66	0.36	0.62	NA	1	1	NA	1
-77>C	(n = 258)	(n = 144)	(n = 92)	(n = 160)	(n = 84)	(n = 78)	(n = 103)	(n = 120)
T	75.8	81.9	81.0	66.3	86.3	97.4	88.3	91.3
C	24.2	18.1	19.0	33.7	13.7	2.6	11.7	8.7
T/T	57.0	68.7	66.3	45.6	73.8	94.9	77.7	83.3
T/C	37.6	26.4	29.3	41.3	25.0	5.1	21.4	15.8
C/C	5.4	4.9	4.3	13.1	1.2	0.0	1.0	0.8
p-HW	0.87	0.25	0.73	0.37	1	1	1	1
5C>A	(n = 262)	(n = 151)	(n = 91)	(n = 161)	(n = 84)	(n = 83)	(n = 104)	(n = 124)
C	76.9	83.1	80.8	66.1	86.3	97.6	88.9	91.5
A	23.1	16.9	19.2	33.9	13.7	2.4	11.1	8.5
C/C	58.8	70.2	65.9	45.3	73.8	95.2	78.8	83.9
C/A	36.3	25.8	29.7	41.6	25.0	4.8	20.2	15.3
A/A	5.0	4.0	4.4	13.0	1.2	0.0	1.0	0.8
p-HW	0.86	0.38	0.73	0.38	1	1	1	1
1057G>A	(n = 274)	(n = 158)	(n = 101)	(n = 161)	(n = 83)	(n = 89)	(n = 110)	(n = 131)
G	78.8	80.4	30.7	92.2	97.0	98.9	85.9	69.1
A	21.2	19.6	69.3	7.8	3.0	1.1	14.1	30.9
G/G	62.0	63.3	8.9	85.7	94.0	97.8	71.8	44.3
G/A	33.6	34.2	43.6	13.0	6.0	2.2	28.2	49.6
A/A	4.4	2.5	47.5	1.2	0.0	0.0	0.0	6.1
p-HW	1	0.45	1	0.24	1	1	0.12	0.10
Mean H	36.9	30.0	33.2	37.1	21.5	6.9	24.3	19.6

AF-B, EU-B, and JA-B, as well as in Australians of European origin [25] and Japanese [17], where the -79°G allele frequency reaches 37%. However, this position is monomorphic in 4 Amerindian populations; that is, the most frequent allele, -79°C, reached frequencies higher than 99%, characterizing monomorphism, including 2 populations (GKW and KRC) where it was fixed. In native Americans, only 4 subjects were heterozygous for this position, 1 GND, 2 GRC, and 1 KIV, and the -79°G allele seems to occur in these populations only because of recent gene flow from non-Amerindian populations. The European and sub-Saharan African gene flow into Amerindian populations is 14% for GND, 4% for GRC, virtually absent in GKW, and 7% for the Kaingang [12,13], as estimated using the highly informative *HLA* alleles as markers. The absence of the

CD80 -79°G allele in the original gene pool of the Amerindian populations can be explained by a founder effect or genetic drift during the period of the migrations from Asia to America or later through the American continent, from North America to South America. The founder individuals did not carry or their descendants did not receive this allele. Also, these populations suffered an extreme decrease in population size after the colonization of the American continent by Europeans, and it is possible that some alleles were fixed as a result of this bottleneck effect.

CD86 c.1057G>A is a nonsynonymous SNP in exon 8, which leads to an alanine-to-threonine exchange at amino acid residue 304 (A304T) and introduces a potential phosphorylation site in the cytoplasmic region of the *CD86* molecule [26]. The supposedly

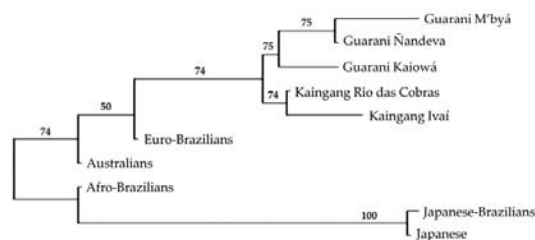


Fig. 1. Dendrogram constructed using the neighbor-joining method and the modified Cavalli-Sforza chord measure of genetic distances for *CD80* promoter variants and *CD86* exon 8 single-nucleotide polymorphism for the populations analyzed in this study as well as Australians of European origin [25] and Japanese [17]. Bootstrap values are indicated above the branches.

ancestral *CD86**G allele is the most common in Europeans and in populations of European ancestry [25–28], as well as in EU-B, AF-B, Africans, and Amerindians. The frequency in EU-B is the highest described for Europeans and populations of European ancestry. This may be caused by the approximately 3 to 7% Amerindian ancestry reported for this population [10,11], since AF-B and EU-B presented similar frequencies and Amerindians had the highest frequencies of this allele (up to 98.9%). The Japanese [17], Chinese [29–31], and JA-B were the only populations in which the 1057*A allele is the most frequent (66, 52.9, 52.5, 53.6, and 69.3%, respectively). Therefore, the significant shift in allele frequencies, with allele 1057*A becoming more frequent than 1057*G, probably occurred recently in Eastern Asian populations. *CD86* allelic frequencies were similar between GRC and GND ($p = 0.21$), KRC and EU-B ($p = 0.10$), and EU-B and AF-B ($p = 0.59$). Also, KIV allelic frequencies were similar to those of French [26], British [27], Finnish [28], and Australian populations of European origin [25]. AF-B and EU-B allelic frequencies were similar to those of the French, and the frequency in JA-B was similar to that of the Japanese [17].

Three of 64 (4.6%) p values for HW were significant ($p = 0.01$, 0.02, and 0.04; Table 2). Deviations from HW equilibrium can be the result of population stratification and sampling effects, but a small number of low-probability p values are expected because of the multiple statistical comparisons performed.

In the neighbor-joining dendrogram based on *CD80* and *CD86* genetic distances (Fig. 1), the Amerindian populations form a cluster, with the Guarani and the Kaingang forming subgroups. The differentiation between the Kaingang and Guarani is consistent with the small gene flow between them estimated using *HLA* haplotypes. In the Guarani gene pool, 1.4% is from Kaingang origin and 0.5% of Kaingang gene pool is derived from the Guarani [12,32]. The most differentiated cluster in the dendrogram was formed by the Japanese and JA-B, which were not clustered with Amerindians. Amerindians are descendants from Eastern Asian populations that migrated through the Bering Strait to North America and subsequently to Central and South America [33,34]. Nevertheless, it is not unexpected that related populations present contrasting allele frequencies at a fraction of the loci. It should be stressed that internal nodes of a dendrogram based on gene frequencies in different

populations of a species do not necessarily represent a common ancestral population and that grouping of populations may result not only from common ancestry but also from gene flow, natural selection, or random genetic drift. The dendrogram reflects the frequencies of the *CD80* and *CD86* alleles and therefore might not reveal population history. To reconstruct the evolutionary history of populations, data from numerous unlinked loci would be required. By contrast, deviations from the expected distribution of populations may reveal the action of different evolutionary factors on the polymorphism or differential magnitude of the same evolutionary factors, especially random genetic drift or natural selection. Considering the unusual distance of the Japanese in our analyses, the change in allele frequency of the *CD80* and *CD86* genes probably occurred in a more recent evolutionary time, in the ancestors of the Eastern Asian populations.

The *CD86* SNP was the most diverse, with heterozygosity values reaching 49.6% in KIV. Within *CD80* polymorphisms, *g*-387T>C had the highest heterozygosity (49.4% in EU-B). Although the African sample was too small for meaningful statistical analysis, with only 16 individuals, all *CD80* and *CD86* alleles were observed (Table 3) and Africans had the highest diversity (average heterozygosity of 43.7%). The presence of all alleles in the African sample suggests that the *CD80* and *CD86* polymorphisms investigated originated in Africa before the human migrations out of that continent. By contrast, the GRC population exhibited the lowest diversity, at 6.9%.

Usually, most genetic diversity is observed within populations [35]. The analysis of molecular variance for the genomic region investigated in this study indicated that 69% of the diversity was observed between Amerindians and urban populations, 2.6% among populations within groups, 0.46% among individuals within populations, and 28.0% within individuals. The interpopulation diversity in Amerindians was 11.4%. This high interpopulation differentiation, especially between Amerindians and urban populations, may reveal the action of evolutionary forces modulating the distribution of the polymorphisms differently in each population.

3.2. Haplotypes

CD80 promoter variations were previously described in only 4 allelic combinations, named *promoter 1* to *promoter 4* [17]. In the present study 5 new *CD80* promoter alleles were described and confirmed by cloning and sequencing. *Promoter 5* (-557_561*del, -454°C, -387*T, -232*G, -79°C, -7°C, 5'A) was present in 1 African individual of mixed African and European ancestry as well as in an AF-B individual. *Promoter 7* (-557_561*del, -454°C, -387*T, -232*G, -79°C, -7°C, 5'C) was described in 6 AF-B. Considering the ancestry of the individuals and the generally higher diversity observed in Africans, *promoter 5* and *promoter 7* are most likely alleles of African origin. *Promoter 6* (-557_561*del, -454°C, -387*T, -232*G, -79°C, -7°C, 5'C) was observed in 2 EU-B and 4 AF-B individuals, and its geographic origin may be Europe or most likely Africa. *Promoter 8* (-557_561*ins, -454*A, -387*T, -232*G, -79°C, -7*T, 5'C) was observed in a KRC and *promoter 9* (-557_561*ins, -454*A, -387*T, -232*A, -79°C, -7*T, 5'C) in a KIV individual, so they may have originated in the American continent. To determine the origins and geographic distributions of these alleles, other Amerindian populations should be investigated.

Table 3
Genotypic frequencies of the *CD80* and *CD86* variations and allelic frequencies of *CD80* promoter alleles in Africans ($n = 16$)

<i>CD80</i> -557_561	<i>CD80</i> -454	<i>CD80</i> -387	<i>CD80</i> -232	<i>CD80</i> -79	<i>CD80</i> -7	<i>CD80</i> 5	<i>CD80</i> alleles	<i>CD86</i> 1057
del/del 8	C/C 8	T/T 11	G/G 8	C/C 10	T/T 8	C/C 8	<i>Promoter 1</i> 7	G/G 8
ins/del 8	C/A 8	T/C 5	G/A 8	C/G 5	T/C 7	C/A 7	<i>Promoter 2</i> 5	G/A 8
ins/ins 0	A/A 0	C/C 0	A/A 0	G/G 1	C/C 1	A/A 1	<i>Promoter 3</i> 11	A/A 0
							<i>Promoter 4</i> 8	
							<i>Promoter 5</i> 1	

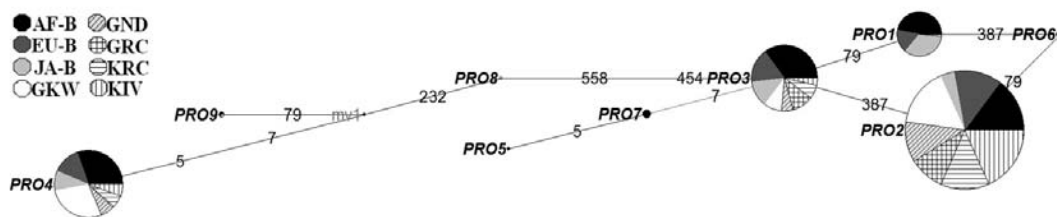


Fig. 2. Median-joining network indicating the relationships among *CD80* promoter (*PRO1–9*) alleles. Cycle size is proportional to the allele relative frequency. The number of copies in each population is indicated for each allele. The numbers correspond to the nucleotides changed. AF-B, Afro-Brazilians; EU-B, Euro-Brazilians; JA-B, Japanese Brazilians; GKW, Guarani Kaiowá; GND, Guarani Nandeva; GRC, Guarani M'byá; KRC, Kaingang from Rio das Cobras; KIV, Kaingang from Ivaí.

The relationship between *CD80* alleles, not considering recombination, is illustrated in Fig. 2. Promoters 1, 2, and 3 are similar in sequence and promoter 4 is the most differentiated. Considering the whole sample, promoter 2 is the most common allele and promoter 1 the least frequent. In JA-B promoter 1 is the most frequent allele, but in Amerindians it is virtually absent because of the monomorphism of nucleotide -79 (Table 4). *CD80* promoter allele frequencies were similar only between the KIV and KRC ($p = 0.08$) and between the GND and KRC ($p = 0.62$). Comparing *CD80* promoter frequencies with Japanese [17] and Australians of European ancestry [25], the only similarity was between the JA-B and Japanese.

Based on the distribution of *CD80* haplotypes in all populations investigated, the nucleotide changes, and the LD between *CD80* and *CD86*, we propose a possible model for the origin and distribution of the *CD80* promoter alleles. The ancestral haplotype seems to be promoter 2, which has global occurrence and the highest frequencies in all populations, except in the Japanese and JA-B, as well as a nonsignificant LD with *CD86* in AF-B, EU-B, and JA-B. The allele promoter 3 might have been originated by a point mutation, the transition from C to T at position -387. Promoter 3 might have originated promoter 1 by a transversion from C to G at position -79 and promoter 7 by a T-to-C transition at nucleotide -7. Promoter 7, by a transversional mutation from C to A at nucleotide 5, originated promoter 5, and this allele, by an insertion of 5 bp at position -557_–561 and another transversion from C to A at position -454 and a transition from G to A at nucleotide -232, originated promoter 4. The other alleles were probably originated by recombination because reverse mutations are rare: promoter 6 from promoter 2 and 1 between nucleotides -387 and -79; promoter 9 from promoter 4 and 1 between positions -232 and -79; and promoter 8 from a recombination involving either promoter 4 and 2 or promoter 4 and 3, between nucleotides -387 and -232. Because of the lack of the intermediate forms, the sequence of the mutations leading from promoter 5 to promoter 4 cannot be inferred. Three simultaneous mutation events at such close positions are unlikely to occur and

other haplotypes intermediate to promoter 5 and 4, carrying 1 or 2 of these mutations, might have existed in ancestral populations and may still exist in populations not investigated or at low frequency, even in the analyzed populations.

3.3. Linkage disequilibrium

All pairs of SNPs in the *CD80* promoter are in high and significant LD, with $D' \approx 0.9$ and p values ≤ 0.01 for all populations (Table 5). The nucleotides -387 and -79, although in high LD, are not highly correlated with the others, as indicated by the low r^2 values. For the combinations of the 4 frequent *CD80* promoter alleles and the *CD86* c.1057G>A SNP, all 8 possible haplotypes were observed (Table 6). In Amerindians, EU-B, and AF-B, most *CD80* promoter alleles occurred preferentially in combination with *CD86* 1057*G. The exception was for *CD80** promoter 3 in KRC, which occurred more frequently with *CD86* 1057*A. In KIV, although the *CD80** promoter 2 occurred more frequently with *CD86* 1057*G, it also occurred with *CD86* 1057*A at a high frequency. In JA-B, *CD80** promoter 1, *2, and *4 occurred more frequently with *CD86* 1057*A and *CD80** promoter 3 occurred with both *CD86* 1057*A and *G at similar frequencies. *CD80** promoter 2 exhibited the lowest LD with *CD86*, only significant in the GKW and Kaingang populations (Table 7). The Kaingang exhibited absolute and always significant LD. The lack of significance in LD for GND and GRC is probably caused by the low frequencies of *CD86* 1057*A (3.0 and 1.1%). GKW was the only Amerindian population that did not present absolute LD.

Although LD was high among all *CD80* polymorphisms, considering $r^2 \geq 0.8$ as the cutoff, nucleotides -387 and -79 were excluded from the LD block, a pattern observed in all samples. When investigating the association with this genomic region, the polymorphisms at positions -557_–561, -454, -232, -7, and 5 can be considered a single block of LD and any of them can be chosen as a tag SNP representing the diversity of the whole block.

Table 4

Haplotype frequencies (%) observed for the combination of *CD80* promoter variants c.-557_–561 insCATGA (-558 ins), c.-454C>A, c.-387T>C, c.-232G>A, c.-79G>C, c.-7T>C, c.5C>A

Haplotypes	AF-B (n = 245)	EU-B (n = 142)	JA-B (n = 89)	GKW (n = 160)	GND (n = 84)	GRC (n = 76)	KRC (n = 98)	KIV (n = 120)
Promoter 1 (DEL+CTGGTC)	15.5	9.5	30.9	0.0	0.6	1.3	0.0	0.0
Promoter 2 (DEL+CCGCTC)	33.9	52.1	24.2	56.6	73.8	71.7	77.0	85.4
Promoter 3 (DEL+CTGCTC)	24.9	20.4	25.3	9.7	11.9	24.3	11.2	5.4
Promoter 4 (CATGA+ATACCA)	23.3	17.6	19.7	33.7	13.7	2.6	11.2	8.8
Promoter 5 (DEL+CTGCCA)	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Promoter 6 (DEL+CCGGTC)	0.8	0.4	0.0	0.0	0.0	0.0	0.0	0.0
Promoter 7 (DEL+CTGCCC)	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Promoter 8 (CATGA+ATGCTC)	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0
Promoter 9 (CATGA+ATAGTC)	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.4

n, number of individuals; AF-B, Afro-Brazilians; EU-B, Euro-Brazilians; JA-B, Japanese Brazilians; GKW, Guarani Kaiowá; GND, Guarani Nandeva; GRC, Guarani M'byá; KRC, Kaingang from Rio das Cobras; KIV, Kaingang from Ivaí.

Table 5
Values of p , D' , and r^2 for linkage disequilibrium between the polymorphisms of the *CD80* promoter

	p	D'	r^2	p	D'	r^2	p	D'	r^2	p	D'	r^2	p	D'	r^2	p	D'	r^2
	-557,-561,-454			-557,-561,-387			-557,-561,-232			-557,-561,-7			-557,-561,5			-454,-387		
AF-B	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.16	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.91	<10 ⁻⁵	1	0.97	<10 ⁻⁵	1	0.16
EU-B	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.23	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.23	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.23
JA-B	<10 ⁻⁵	1	1	0.0002	1	0.08	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	0.0002	1	0.08
GKW	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.66	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.66
GND	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.45	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.45
GRC	<10 ⁻⁵	1	1	0.0013	1	0.07	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	0.0013	1	0.07
KRC	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.45	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.42	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.45
KIV	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.59	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	0.59
	-454,-232			-454,-7			-454,5			-387,-232			-387,-7			-387,5		
AF-B	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.91	<10 ⁻⁵	1	0.97	<10 ⁻⁵	1	0.16	<10 ⁻⁵	1	0.17	<10 ⁻⁵	1	0.16
EU-B	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.23	<10 ⁻⁵	1	0.23	<10 ⁻⁵	1	0.23
JA-B	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	0.0002	1	0.08	0.0002	1	0.08	0.0002	1	0.08
GKW	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.66	<10 ⁻⁵	1	0.66	<10 ⁻⁵	1	0.66
GND	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.45	<10 ⁻⁵	1	0.45	<10 ⁻⁵	1	0.45
GRC	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	0.0013	1	0.07	0.0013	1	0.07	0.0013	1	0.07
KRC	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	0.42
KIV	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	0.59	<10 ⁻⁵	1	0.56	<10 ⁻⁵	1	0.56
	-232,-7			-232,5			-7,5			-557,-561,-79			-454,-79			-387,-79		
AF-B	<10 ⁻⁵	1	0.91	<10 ⁻⁵	1	0.97	<10 ⁻⁵	1	0.95	<10 ⁻⁵	0.9	0.05	<10 ⁻⁵	0.9	0.05	<10 ⁻⁵	0.9	0.08
EU-B	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	0.0100	1	0.02	0.0100	1	0.02	<10 ⁻⁵	0.9	0.10
JA-B	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	0.11	<10 ⁻⁵	1	0.11	<10 ⁻⁵	1	0.14
GKW	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	NA			NA			NA		
GND	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	NA			NA			NA		
GRC	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	NA			NA			NA		
KRC	<10 ⁻⁵	1	0.42	<10 ⁻⁵	1	1	<10 ⁻⁵	1	1	NA			NA			NA		
KIV	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	0.95	<10 ⁻⁵	1	1	NA			NA			NA		
										-232,-79			-79,-7			-79,5		
AF-B	<10 ⁻⁵			<10 ⁻⁵			<10 ⁻⁵			<10 ⁻⁵	0.9	0.05	<10 ⁻⁵	1	0.07	<10 ⁻⁵	1	0.06
EU-B										0.0100	1	0.02	0.0100	1	0.02	0.0100	1	0.02
JA-B										<10 ⁻⁵	1	0.11	<10 ⁻⁵	1	0.11	<10 ⁻⁵	1	0.11

NA, not analyzed; AF-B, Afro-Brazilians; EU-B, Euro-Brazilians; JA-B, Japanese Brazilians; GKW, Guarani Kaiowá; GND, Guarani Nandeva; GRC, Guarani M'byá; KRC, Kaingang from Rio das Cobras; KIV, Kaingang from Ivaí.

The functional relevance of the *CD80* polymorphisms analyzed in this study is not yet known. The *CD80* polymorphisms are all located in the promoter region and the 5' untranslated region of the

gene, so they can alter gene expression, most likely through differential binding of factors regulating the transcription or the translation. The *CD86* c.1057G>A SNP is expected to create a phosphory-

Table 6
Haplotypic frequencies (%) observed for *CD80* and *CD86* polymorphisms

<i>CD80/CD86</i>	AF-B (n = 241)	EU-B (n = 141)	JA-B (n = 88)	GKW (n = 160)	GND (n = 83)	GRC (n = 74)	KRC (n = 98)	KIV (n = 120)
*pro1/1057*A	1.2	2.5	28.4	0.0	0.0	0.0	0.0	0.0
*pro1/1057*G	14.7	7.1	2.3	0.0	0.6	1.4	0.0	0.0
*pro2/1057*A	8.3	9.2	17.0	6.9	3.0	1.4	0.0	31.7
*pro2/1057*G	25.3	42.9	7.4	49.7	71.1	71.6	62.7	53.7
*pro3/1057*A	2.7	7.8	11.9	0.3	0.0	0.0	14.3	0.0
*pro3/1057*G	21.8	12.8	13.6	9.4	12.0	23.6	11.2	5.4
*pro4/1057*A	9.3	0.7	13.6	0.6	0.0	0.0	0.0	0.0
*pro4/1057*G	14.1	17.0	5.7	33.1	13.2	2.0	11.2	8.7
Others	2.5	0.0	0.0	0.0	0.0	0.0	0.5	0.4

n, number of individuals; AF-B, Afro-Brazilians; EU-B, Euro-Brazilians; JA-B, Japanese Brazilians; GKW, Guarani Kaiowá; GND, Guarani Nandeva; GRC, Guarani M'byá; KRC, Kaingang from Rio das Cobras; KIV, Kaingang from Ivaí.

Table 7
Values of p , D' , and r^2 for linkage disequilibrium between the *CD80* and *CD86* genes

	<i>CD80 Pro 1, CD86</i>			<i>CD80 Pro 2, CD86</i>			<i>CD80 Pro 3, CD86</i>			<i>CD80 Pro 4, CD86</i>		
	p	D'	r^2	p	D'	r^2	p	D'	r^2	p	D'	r^2
AF-B	<0.01	0.65	0.022	0.31	0.06	0.002	<0.01	0.50	0.0228	<0.01	0.25	0.0568
EU-B	0.44	0.07	0.002	0.27	0.12	0.004	<0.01	0.23	0.0504	<0.01	0.80	0.0351
JA-B	<0.01	0.74	0.100	0.84	0.02	<0.001	<0.01	0.34	0.0990	0.95	0.01	0.0000
GKW	NA	NA	NA	<0.01	0.72	0.034	0.32	0.59	0.0031	<0.01	0.76	0.0251
GND	NA	NA	NA	0.18	1	0.011	0.40	1	0.0043	0.37	1	0.0047
GRC	NA	NA	NA	0.39	1	0.005	0.43	1	0.0042	0.84	1	0.0003
KRC	NA	NA	NA	<0.01	1	0.281	<0.01	1	0.4867	0.04	1	0.0211
KIV	NA	NA	NA	<0.01	1	0.079	0.01	1	0.0265	<0.01	1	0.0444

NA, not analyzed; AF-B, Afro-Brazilians; EU-B, Euro-Brazilians; JA-B, Japanese Brazilians; GKW, Guarani Kaiowá; GND, Guarani Nandeva; GRC, Guarani M'byá; KRC, Kaingang from Rio das Cobras; KIV, Kaingang from Ivaí.

lation site in the cytoplasmic region of the CD86 molecule, so it can potentially interfere in the signal transduction on APCs and in the regulation of T-cell activation.

4. Conclusions

CD80 promoter and CD86 exon 8 allele frequencies vary significantly among populations of different ancestries, and all polymorphisms originated in Africa before the human migrations out of that continent. The 5 new CD80 promoter alleles described are likely of African, European, and Amerindian origin, and the proposed model of origin of the alleles indicates that promoter 2 is most likely the ancestral allele. The distribution of allele frequencies in CD86 1057 SNP was in accordance with the previously reported populations of similar ancestry, with Japanese and Chinese exhibiting opposite patterns compared with other populations, including Amerindians. In most cases, CD80 and CD86 follow the overall pattern described in several other loci, with Amerindians having lower genetic diversity and more pronounced LD than other populations and even being monomorphic for nucleotide -79 of the CD80 promoter. However, the GWK presented high heterozygosity for most CD80 polymorphisms and the KIV for CD86 SNP. LD analyses suggested that 5 of the 7 polymorphisms analyzed in the CD80 promoter are highly correlated and can be considered an LD block.

In this study we analyzed individuals of European, African, Japanese, and Amerindian origin in addition to the other European, Japanese, and Chinese samples previously described in the literature. For the complete comprehension of CD80 and CD86 polymorphism evolution and diversity, other populations remain to be investigated worldwide.

Acknowledgments

We are grateful for the cooperation of all individuals of the studied populations and appreciate the comments and technical assistance of our laboratory colleagues. This work was supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Institutos do Milênio, PRONEX, Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Paraná, and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). The funding sources had no role in the study design, collection, analysis, and interpretation of the data, or the writing or decision to submit the manuscript for publication.

References

- Keir ME, Sharpe AH. The B7/CD28 costimulatory family in autoimmunity. *Immunol Rev* 2005;204:128–43.
- Chambers CA. The expanding world of co-stimulation: the two-signal model revisited. *Trends Immunol* 2001;22:217–23.
- Collins AV, Brodie DW, Gilbert RJC, Iaboni A, Manso-Sancho R, Walse B, et al. The interaction properties of costimulatory molecules revisited. *Immunity* 2002;17:201–10.
- Bhatia S, Edidin M, Almo SC, Nathanson SG. B7-1 and B7-2: similar costimulatory ligands with different biochemical, oligomeric and signaling properties. *Immunol Lett* 2006;104:70–5.
- Sansom DM, Manzotti CM, Zheng Y. What's the difference between CD80 and CD86? *Trends Immunol* 2003;24:313–8.
- Suvas S, Singh V, Sahdev S, Vohra H, Agrewala JN. Distinct role of CD80 and CD86 in the regulation of the activation of B cell and B cell lymphoma. *J Biol Chem* 2002;277:7766–75.
- Lonjou C, Zhang W, Collins A, Tapper WJ, Elahi E, Maniatis N, et al. Linkage disequilibrium in human populations. *Proc Natl Acad Sci USA* 2003;100:6069–74.
- Hey J. On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biol* 2005;3:e193.
- Faucz FR, Probst CM, Petzl-Erler ML. Polymorphism of LMP2, TAP1, LMP7 and TAP2 in Brazilian Amerindians and Caucasoids: implications for the evolution of allelic and haplotypic diversity. *Eur J Immunogenet* 2000;27:5–16.
- Braun-Prado K, Vieira Mion AL, Farah Pereira N, Culp L, Petzl-Erler ML. HLA class I polymorphism, as characterised by PCR-SSOP, in a Brazilian exogamic population. *Tissue Antigens* 2000;56:417–27.
- Probst CM, Bompeixe EP, Pereira NF, de O Dalajio MM, Visentainer JEL, Tsuneto LT, et al. HLA polymorphism and evaluation of European, African, and Amerindian contribution to the white and mulatto populations from Paraná, Brazil. *Hum Biol* 2000;72:597–617.
- Petzl-Erler ML, Luz R, Sotomaior VS. The HLA polymorphism of two distinctive South-American Indian tribes: the Kaingang and the Guarani. *Tissue Antigens* 1993;41:227–37.
- Tsuneto LT, Probst CM, Hutz MH, Salzano FM, Rodriguez-Delfino LA, Zago MA, et al. HLA class II diversity in seven Amerindian populations. Clues about the origins of the Aché. *Tissue Antigens* 2003;62:512–26.
- Lahiri DK, Nurnberger J Jr. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res* 1991;19:5444.
- Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning. A Laboratory Manual*. 2d ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.
- Bignon JD, Fernandez-Viña MA. Protocols of the 12th IHW for PCR-SSOP. In: Charon D, editor. *Genetic Diversity of HLA: Functional and Medical Implications*. Paris: EDK; 1997, pp 584–95.
- Matsushita M, Tsuchiya N, Oka T, Yamane A, Tokunaga K. New polymorphisms of human CD80 and CD86: lack of association with rheumatoid arthritis and systemic lupus erythematosus. *Genes Immun* 2000;1:428–34.
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999;41:95–8.
- Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 1992;48:361–72.
- Miller MP. RxC: a program for the analyses of contingency tables via the Metropolis Algorithm computer. In: *Computer Software Distributed by Author* 1997.
- Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 2005;1:47–50.
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–25.
- Nei M, Tajima F, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* 1983;19:153–70.
- Ota T. DISPAN: Genetic Distance and Phylogenetic Analysis. University Park PA: Institute of Molecular Evolutionary Genetics, Pennsylvania State University; 1993.
- Teutsch SM, Booth DR, Bennetts BH, Heard RNS, Stewart GJ. Association of common T cell activation gene polymorphisms with multiple sclerosis in Australian patients. *J Neuroimmunol* 2004;148:218–30.
- Delneste Y, Bosotti R, Magistrelli G, Bonnefoy JY, Gauchat JF. Detection of a polymorphism in exon 8 of the human CD86 gene. *Immunogenetics* 2000;51:762–3.
- Abdallah AM, Renzoni EA, Anevlavis S, Lagan AL, Munkonge FM, Fonseca TC, et al. A polymorphism in the promoter region of the CD86 (B7.2) gene is associated with systemic sclerosis. *Int J Immunogenet* 2006;33:155–61.
- Turpeinen H, Laine A-P, Nejentsev S, Sjöröos M, Ilonen J, Simell O, et al. CD86 Gene polymorphisms: no association with type 1 diabetes among Finnish subjects. *Diabetologia* 2002;45:1041–2.
- Liu Y, Liang W-B, Gao L-B, Pan X-M, Chen T-Y, Wang Y-Y, et al. CTLA4 and CD86 gene polymorphisms and susceptibility to chronic obstructive pulmonary disease. *Hum Immunol* 2010;71:1141–6.
- Ma X-N, Wang X, Yan Y-Y, Yang L, Zhang D-L, Sheng X, et al. Absence of association between CD86 +1057G/A polymorphism and coronary artery disease. *DNA Cell Biol* 2010;29:325–8.
- Pan X-M, Gao L-B, Liang W-B, Liu Y, Zhu Y, Tang M, et al. CD86 +1057 G/A polymorphism and the risk of colorectal cancer. *DNA Cell Biol* 2010;29:381–6.
- Oliveira LA, Ribas F, Bicalho MG, Tsuneto LT, Petzl-Erler ML. High frequencies of alleles MICA*020 and MICA*027 in Amerindians and evidence of positive selection on exon 3. *Genes Immun* 2008;9:697–705.
- Rogers RA, Rogers LA, Martin LD. How the door opened: the peopling of the New World. *Hum Biol* 1992;64:281–302.
- Wallace DC, Torroni A. American Indian prehistory as written in the mitochondrial DNA: a review. *Hum Biol* 1992;64:403–16.
- Barbujani G, Magagnoli A, Minch E, Cavalli-Sforza LL. An apportionment of human DNA diversity. *Proc Natl Acad Sci USA* 1997;94:4516–9.

4 CAPÍTULO II

Polymorphisms in the promoter region of the *CD80* and *CD86* genes alter gene expression in vitro

Márcia Holsbach Beltrame^a, Arthur A Simen^b, Ricardo Dalla-Costa^{a,1}, Becky C Carlyle^b, Juarez Gabardo^a, Albert C Shaw^b, Maria Luiza Petzl-Erler^{a*}

^aHuman Molecular Genetics Laboratory, Department of Genetics, Federal University of Paraná

P.O. Box 19071

81531-980 Curitiba, Brazil

Márcia Holsbach Beltrame: marcia.hbeltrame@gmail.com

Ricardo Dalla-Costa: dallacosta_r@yahoo.com.br

Juarez Gabardo: jgabardo@ufpr.br

^bDepartment of Psychiatry, Yale School of Medicine, New Haven, CT, USA

Arthur A Simen: arthur.simen@yale.edu

Becky C Carlyle: becky.carlyle@gmail.com

Albert C Shaw: albert.shaw@yale.edu

*Corresponding autor.

E-mail address: perler@ufpr.br

Tel. (+55 41) 33611746

Fax (+55 41) 33611793

¹Present address: Life Technologies, Avenida do Café, 277, 1º Andar, Torre A. Bairro Vila Guarani, 04311-000 São Paulo, Brazil.

ABSTRACT

The *CD80* and *CD86* genes, also called *B7.1* and *B7.2*, are located on chromosome 3 and code for glycoproteins of the immunoglobulin superfamily that are expressed on the surface of antigen presenting cells. As co-stimulatory molecules, CD80 and CD86 play essential roles in the regulation of T cell responses through binding to the CD28 and CTLA-4 receptors. Here, the functional relevance of *CD80* and *CD86* promoter polymorphisms in promoter activity was investigated using luciferase reporter assays in HEK293T cells, flow cytometry measurements of protein on TLR stimulated monocytes of individuals with different genotypes, and *in silico* inferences of transcription factor binding. The *CD80* and *CD86* promoter regions were sequenced in individuals of various ancestries and within the proximal 2.3 kb of the *CD80* promoter and 5'UTR, 23 polymorphisms were found in four common haplotypes. In *CD86*, two SNPs were analyzed in the 2.4 kb including the promoter and 5'UTR of the gene. Luciferase measurements showed that the *CD86* -298T>C SNP did not interfere in gene expression in the conditions tested. The *CD80* haplotype 4 showed significantly lower promoter activity. Using seven recombinant constructs produced with different segments of haplotypes 2 and 4, we found that SNPs *g*.-7T>C (rs16829980), *g*.5C>A (rs41271391) and/or 287A>T (rs56124423) are responsible for the decreased expression. The SNP *g*.-7T>C is located in the previously reported binding site of activating protein 1 (AP-1) and is predicted to interfere with AP-1 binding. The protein expression on monocytes differed significantly among *CD80* haplotypes and *CD86* -819T>C (rs11575855) alleles, and also between age groups and the different stimuli given to the cells. We conclude that polymorphisms within the 3' end of the *CD80* promoter significantly affect the activity of the promoter. Further, the *CD86* -819T>C SNP has a significant effect on protein levels on the membrane of stimulated monocytes. These differences are most likely caused by differential binding of transcription factors.

Keywords: *CD80*, *CD86*, gene expression, reporter gene assay, genetic polymorphism, haplotype

1. Introduction

CD80 and *CD86* are linked genes on chromosome 3q21 that code for costimulatory molecules expressed on the surface of antigen presenting cells (APC), such as dendritic cells, monocytes and activated B cells. The CD80 and CD86 molecules, also called B7.1 and B7.2, participate of the regulation of T cell activation through binding to the CD28 and CTLA-4 coreceptors (Chambers, 2001). The costimulatory signal provided by B7 and CD28 interaction is essential for full activation, proliferation and differentiation of T lymphocytes into effector cells after the antigen specific signal derived by the interaction of the MHC (major histocompatibility complex) molecules/peptide complex with the T cell receptor (TCR). The interaction of B7 and CTLA-4 (CD152), a receptor induced after T cell stimulation that interacts with the B7 ligands with higher affinity than CD28, provides a negative signal which controls the proliferative response, having a critical role in the prevention of autoimmune disease (Keir and Sharpe, 2005). Although CD80 and CD86 are both able to costimulate T cell proliferation, they have different expression patterns. CD86 is constitutively expressed with maximum expression between 24 and 48 hours after interaction with CD28, while CD80 is upregulated upon T cell activation and stable for 4 to 5 days. CD86 binds CD28 more efficiently than CD80 and is its preferential ligand, while CTLA-4 has higher affinity for CD80 than for CD86 (Bhatia et al., 2006; Samson, Manzotti and Zheng, 2003; Collins et al., 2002 and Suvas et al., 2002).

Due to their important role in immune responses, *CD80* and *CD86* expression was investigated in patients with autoimmune diseases and allergy conditions. The soluble forms of CD80 e CD86 were significantly increased in the plasma of patients with systemic lupus erythematosus compared to control individuals. Plasma concentrations of CD80 and CD86 were also directly correlated with the severity of acute asthma in children (Wong et al., 2005a; Wong et al., 2005b; Ip et al., 2006). Increased levels of CD80 molecule on the surface of stimulated monocytes was also found in individuals presenting a better response to influenza vaccination (Duin et al., 2007).

Polymorphisms in the promoter region of a gene can affect transcription by changing the affinity of the DNA sequence for regulatory proteins in the nucleus. This occurs because the ligation of transcription factors that can activate or inhibit transcription through the formation of complex protein-DNA and protein-protein interactions depends on the nucleotide sequence. Thus, alleles can differ in their affinity for regulatory proteins resulting in quantitative variation of the messenger RNA, with functional consequences.

The promoter region of *CD80* and *CD86* is not fully characterized but there are some evidences of functional regions. Fong, Wu and Kipps (1996) identified a positive regulatory element in the promoter region of *CD80* named B7-RE, between nucleotides -60 and -47. They also described an AP-1 binding motif in a sequence that was later identified as containing a substitution of C for T at position -7 that is predicted to disrupt the binding site by TFSEACH (Matsushita et al., 2000). Lim et al. (2005) found an IFN 7 (IRF-7) binding site between positions -84 and -72 and showed that mutations in this sequence stop transcription.

The *CD86* -359delAAG, located in the 5'UTR, upstream from the exon 2, according to TRANSFAC software (Wingender et al., 2000) is not located in any transcription factor binding site (TFBS) (Teutsch et al., 2004). TFBS were predicted in the *CD86* gene using Transfast (Gesellschaft für Biotechnologische Forschung, Braunschweig, Germany), including at least two NF- κ B binding sites located at -612 (5' GGGACATCTA 3') and -238 (5' GGGAATTTC 3') (Li et al., 1999).

The aim of this study was to identify the functional consequences of common polymorphisms located in the first 2 kb in the promoter region of the *CD80* and *CD86* genes in gene transcription.

2. Materials and Methods

2.1 Sequencing of *CD80* and *CD86* promoters and evolutionary analysis

Genomic DNA was extracted from peripheral blood cells by salting-out (Lahiri and Nurnberg, 1991) and phenol-chloroform-isoamyl-alcohol (Sambrook et al., 1989) methods. Individuals were divided in groups according to their ancestry: 25 Euro-Brazilians (EU-B, of predominantly European ancestry), 15 Brazilians of Japanese ancestry (JA-B), 9 Amerindians of the Guarani M'byá (GRC), Guarani Kaiowá (GKW), Guarani Ñandeva (GND) and Kaingang (KRC and KIV) populations from Paraná and Mato Grosso do Sul states, in Brazil, and 9 Africans (AFRI), mostly from Guinea-Bissau and Angola, and also from Cameroon and Ghana. All individuals signed an informed consent before donating their samples.

A fragment of 2489 bp from the *CD80* promoter and 2483 bp from the *CD86* promoter were amplified by polymerase chain reaction (PCR) and quantified in a 1% agarose gel. PCR products were then purified using exonuclease I (EXO1, USB, Cleveland, OH, USA) and shrimp alkaline phosphatase (SAP, USB, Cleveland, OH, USA) enzymes and the eight sequencing reactions performed for each gene used DYEnamic ET Terminator Cycle Sequencing Kit (GE Healthcare, Pollards Wood, UK). After purification the reactions were injected in the MegaBACE 1000 sequencer (GE Healthcare, Pollards Wood, UK). Sequences

were visually analyzed and contigs were made using Phred/Phrap (Ewing et al., 1998) and CONSED (Gordon et al., 1998). Sequences were then aligned using Clustal W (Thompson, Higgins and Gibson, 1994) with the BioEdit software (Hall, 1999). Good quality sequences were visually inspected for polymorphisms and haplotype determination. Ewens-Watterson neutrality test was performed for *CD80* haplotype frequencies with the Arlequin 3.5.1.2 software package (Excoffier and Lisher, 2010).

2.2 Construction of luciferase expression vectors

The promoter regions of the *CD80* and *CD86* genes were amplified in separate PCRs using genomic DNA from homozygote individuals for each *CD80* and *CD86* haplotype, confirmed by sequencing. For *CD86* primers were *CD86_FOR* 5' GTGTCTCTCTGGGAGGGAGC 3' and *CD86_REV5*' TGAAGGACACACATCAAAGCACCAT 3' for the first PCR and *CD86_FOR2* 5' GAGCACAGGCTGGGGTTTCT 3' and *CD86_REV2* 5' GTGAAAGGCTGAGGTTGAAGATGGT 3' for nested PCR. The 2462 bp fragment, including the *CD86* promoter and exon 1, included the SNPs *-1934A>G* (rs11712647) and *-298T>C* (rs79111254). The PCR contained 100 ng genomic DNA, 0.2 μ M of each primer, 2.5 U of Pfu Ultra II Fusion HS DNA polymerase (Stratagene, La Jolla, CA, USA), 1X Pfu Ultra buffer containing 2 mM MgCl₂, 0.25 mM dNTP mixture, in a total volume of 50 μ L. In the second PCR, instead of genomic DNA we used 3.5 μ L of the PCR product from the first PCR. Reaction started with an initial denaturation step of 95°C for 2 min followed by 30 cycles of denaturation at 95°C for 20 sec, annealing at 63°C in the first PCR and 58°C in the nested PCR for 20 sec and extension at 72°C for 45 sec and a final extension step at 72°C for 3 min.

For *CD80*, primers were designed including a sequence to be cut by the *KpnI* and *SacI* restriction enzymes (underlined) *CD80_FOR*: 5' GGTACCTCGGGTTCGGCTTCCTTTGCCC 3', *CD80_REV*: 5' GAGCTCTGGGAAGCCACAGCCTCCTGTG 3', resulting in a 2359 bp fragment. The PCR mixture contained 100 ng of genomic DNA, 0.2 μ M of each primer, 2.5 U of Pfu Ultra II Fusion HS DNA polymerase (Stratagene, La Jolla, CA, USA), 1X Pfu Ultra buffer containing 2 mM MgCl₂, 0.25 mM dNTP mixture and 5% DMSO (dimethyl sulfoxide) in a total volume of 50 μ L. Reaction starts in an initial denaturation step of 95°C for 2 min followed by 30 cycles of denaturation at 95°C for 20 sec, annealing and extension at 69°C for 1min and 30 seconds and a final extension step at 72°C for 3 min.

PCR products were gel purified and subcloned into pSmart-HCAmp plasmid vector (Lucigen Corporation, Middleton, WI, USA) according to the manufacturer's protocol. The plasmid containing the insert was transformed into DH5 α competent *Escherichia coli* cells (Invitrogen Corporation, Carlsbad, CA, USA) and clones were confirmed by sequencing. The inserts were excised from the vector using *KpnI* and *SacI* enzymes (New England BioLabs, Ipswich, MA, USA) and cloned into pGL4.18 luc2P/Neo vector (Promega, Madison, WI, USA) pre-cut with the same enzymes, upstream from the firefly luciferase promoterless gene. After another transformation, clones were sequenced to confirm that no mutation was introduced during the procedures.

A mutated haplotype was created by deleting an insertion present on *CD80* haplotype 4, -557_-561insCATGA, using PCR. Two fragments were amplified by PCR, using either the forward or the reverse primer and a primer located over the polymorphism, representing the deletion allele (FOR: 5' ATGCAAGGTGCTGTCAGGTTCCACCTGTGCAAACAAGTACAC 3' and REV: 5' TTTCCAAGTACTAGGTTGCCACCCACCCCTGTGTGTCCCTTT 3'). Then both PCR products were used as a template in a PCR containing only the external forward and reverse primers, in order to produce the whole fragment. To produce recombinant constructs (Figure 1) *CD80* haplotypes 2 and 4 were cut with *AfeI*, *HindIII* and *AvrII* enzymes, fragments were gel purified and the new combinations were ligated again into the pGL4.18 vector using T4 DNA ligase (New England BioLabs, Ipswich, MA, USA). All sequences were confirmed by sequencing the recombinant constructs.

2.3 Cell culture

HEK293T cells (from American Type Culture Collection, ATCC, Manassas, VA, USA) were cultured in Dulbecco's modified Eagle medium (DMEM) (GIBCO Invitrogen cell culture, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum, 2 U/ml penicillin, 2 μ g/ml streptomycin and 2 mmol/l L-glutamine (GIBCO Invitrogen cell culture, Carlsbad, CA, USA) and grown in a humidified incubator at 37°C and 5% CO₂.

2.4 Transient transfection and luciferase reporter assay

HEK293T cells, a human cell line derived from embryonic kidney, were grown for 24 hours in 48-well cell culture plates (Becton Dickinson, Franklin Lakes, NJ, USA), coated with PEI (polyethylenimine 800kD) (Sigma, St. Louis, MO, USA) at a density of 3.5x10⁵ cells per well, and then transiently transfected with pGL4.18 vector containing one of the *CD80* or

CD86 haplotypes, the vector without insert (negative control) or pGL3-pro (positive control), using Effectene Transfection Reagent (Qiagen Inc., Duesseldorf, Germany) according to manufacturer's instructions. Cells were harvested 48 hours after transfection, when they were washed three times with 1X PBS and then rocked for 15 min in 1X Passive Lysis Buffer (Promega, Madison, WI, USA). Samples were then assayed for luciferase expression using Luciferase Assay Kit (Promega, Madison, WI, USA), with 10 μ L of cell lysate and 50 μ L of luciferin in a white opaque 96-well flat-bottom microplate (Becton Dickinson, Franklin Lakes, NJ, USA), read in a DTX 880 Series Multimode Detector (Beckman Coulter, Fullerton, CA, USA). Assays were performed in triplicates, in at least three different experiments. Significance was tested by one-way ANOVA and Tukey test, correcting for experiment day.

2.5 *In silico* analysis

A sequence of 40 bp around each polymorphism was tested against a transcription factor database (TRANSFAC) using TESS software (*Transcription Element Search Software*, available online at <http://www.cbil.upenn.edu/cgi-bin/tess/tess>), to determine if it is located in a potential TFBS.

To evaluate the phylogenetic conservation of the sequences, human *CD80* and *CD86* promoters were aligned with the chimpanzee, rhesus macaque and marmoset sequences, obtained from NCBI and aligned in Mulan website (<http://mulan.dcode.org/>), using TBA software (threaded blockset aligner). Alignments were visually inspected to identify conserved regions.

2.6 Genotyping and flow cytometry analysis of the protein expression

Sixty individuals, thirty young (21 to 30 years of age) and thirty elderly (≥ 65 years) had their protein levels at the cell membrane correlated with their respective *CD80* and *CD86* promoter genotypes, determined by MassARRAY system (Sequenom) in the Yale Center for Genome Analysis, Keck Biotechnology Resource Laboratory. These individuals were part of the sample analyzed by Duin and colleagues (2007), when they quantified CD80 and CD86 molecules by flow cytometry on the cell membrane of stimulated monocytes. Cells were stimulated with selected toll-like receptor (TLR) ligands: PAM3C (N-palmitoylS-bis palmitoyloxy propyl cystein), LTA (lipoteichoic acid), R-848 (imidoazoquinoline) and CPG (CpG containing DNA). Measurements were done before and after the stimulus and values were represented as the percentage of increase in protein amounts.

CD80 haplotypes were inferred from the genotypes of three SNPs using the linkage disequilibrium information. Regarding the variables haplotype, age and type of stimulus, one-way analysis of variance (ANOVA) was applied considering completely randomized design and factorial scheme, to consider age and stimulus variation. F test was applied and when necessary the Tukey test for averages. Box plot graphics were done in GraphPad Prism 5.04.

3. Results and Discussion

3.1 New polymorphisms and evolutionary analysis

Within the proximal 2.3 kb of the *CD80* promoter and the 5'UTR, 23 polymorphisms were found, one of them being new, -757_-761insTTTT, an insertion of 4 thymines after a repetition of 20 thymines, which occurs in haplotype 4.

Four haplotypes were described in Japanese (Matsushita et al., 2000) considering six SNPs, c.5A>C, located in the 5'UTR (untranslated region), and the others in the promoter region, c.-454A>C, c.-387T>C, c.-232A>G, c.-79G>C e c.-7C>T, and one indel (insertion/deletion polymorphism), c.-557_-561insCATGA, and it was suggested that these haplotypes may be under balancing selection due to their high frequency in Japanese population. The same haplotypes were found in Australians of European descent (Teutsch et al., 2004). In a previous study of our group, the analysis of Africans, Guarani and Kaingang Amerindians and Brazilians of Japanese, of predominantly European and of mixed African and European ancestry, showed that frequencies varied significantly among populations and five new haplotypes - or *CD80* promoter alleles - were identified (Beltrame et al., 2011). All the 23 polymorphisms found in the present study are also occurring in the same 4 common haplotypes, although a few individuals did not present all of the variants in haplotype 4, the most differentiated haplotype (Table 1).

The Ewens-Watterson neutrality test revealed that these haplotypes may be under balancing selection in Japanese descendents only (Table 2). Their alleles have intermediate frequencies and the observed F value (0.25641) was less than the expected value (0.61065), indicating balancing selection or a recent bottleneck effect or population structure. It is worth noting that our Brazilians of Japanese ancestry and the Japanese population (Matsushita et al., 2000) have similar *CD80* promoter frequencies, so this finding represents an effect occurring in the Japanese population rather than in our populations sample of Japanese descent. In the other populations analyzed, frequencies were not well distributed among the haplotypes, with haplotype 2 being the most frequent especially in Amerindians and haplotype 1 the most rare, even absent in some Amerindian populations, and also five rare new haplotypes.

3.2 *In silico* analysis of transcription factor binding sites (TFBS)

The sequences including the polymorphisms were investigated for differential binding of transcription factors *in silico* (Table 3). In *CD80* promoter all except two SNPs were located in predicted binding sites. Seven alleles disrupted the binding sites and four created new ones. The other ten lead to the binding of different transcription factors. In *CD86* promoter one SNP disrupted the binding sites and the other created new ones.

The human sequence was aligned with the sequences of chimpanzee, rhesus macaque and marmoset and the conservation around the polymorphism positions was evaluated (some examples are shown in Figure 2). The phylogenetic conservation is considered an indicative of functionality, although functional sequences do not need to be conserved. In our analysis, the functional polymorphisms suggested by luciferase experiments were not located in sequences highly conserved among species. Even when the sequence has an important regulatory function, sometimes conservation is not observed because other species may have another TFBS that binds by another regulatory protein with a similar function. Other possibility is that the function developed by this sequence is not essential in other species, and in this case conservation is unlikely.

3.3 *CD80* and *CD86* promoter activity in the luciferase reporter assay

Each haplotype was used to drive luciferase expression in reporter assays in HEK293T cells. Luciferase measurements showed that *CD80* haplotypes 1, 2 and 3 have similar activity, while haplotype 4 has significantly lower promoter activity (Figure 3A). In order to identify the polymorphisms responsible for this difference, seven recombinant haplotypes were produced with different combinations of haplotypes 2 and 4 (Figure 1). Only two recombinant haplotypes had luciferase measurements similar to haplotype 4, and both contained the 3' end of the sequence from haplotype 4, including only 3 SNPs in one of them (Figure 3B and C). These results indicate that SNPs *g.-7T>C* (rs16829980), *g.5C>A* (rs41271391) and/or *288A>T* (rs56124423) are responsible for the decreased expression of haplotype 4 compared to the other three haplotypes. The SNP *g.-7T>C* is located in the previously reported binding site (Fong et al., 1996) of activating protein 1 (AP-1) and is predicted to interfere with AP-1 binding. In order to better understand the regulation of *CD80* expression it will be necessary to test each of these polymorphisms individually and confirm the findings *ex vivo*, as here they were used to drive luciferase activity in one cell type, and results can be different for other cells. Also, *in vitro* conditions do not mimic physiological environment.

For *CD86*, one SNP was analyzed in the 2.4 kb segment including the promoter and the 5'UTR of the gene. In luciferase experiments *CD86* -298*T and -298*C alleles did not differ in gene expression (Figure 3D). According to the TFBS prediction, this polymorphism created new regulatory sites, but even though the sequence is predicted to bind a regulatory factor, the proteins may not be available in the specific cell types analyzed, at least at the moment of gene expression, consequently not participating in gene regulation.

3.4 Expression of the CD80 and CD86 proteins on stimulated monocytes by flow cytometry analysis

The increase of CD80 and CD86 protein expression in TLR induced monocytes was compared among individuals with different promoter alleles. All comparisons were performed when at least two individuals could be included in each genotype. Rare genotypes were grouped with the heterozygous individuals, as -7 C, that includes individuals with both C/C and C/T genotypes. The young group showed significantly higher expression compared to the elderly (Figure 4A and B). The four different stimuli applied to the cells also had different effects on gene expression, with PAM3C being the most effective one (Figure 4C and D).

All the polymorphisms genotyped presented significant influence on gene expression (Figures 4E-H). The same applies to the *CD80* haplotypes (Figures 4I-Q). Haplotype 1 had the highest increase in CD80 expression in relation to unstimulated cells, followed by haplotypes 4, 2 and 3. For the *CD86* gene, SNP -298T>C was not analyzed since all individuals had genotype T/T. For *CD86* -1930A>G genotype AA showed a significantly higher increase in gene expression in comparison to the T+ genotypes (A/T and T/T).

It is interesting to note that while *CD80* haplotype 4 had the lowest expression in luciferase experiments (in which cells were not stimulated), when monocytes were stimulated the increase in expression was higher than that of other haplotypes, showing a better potential to respond to the stimulus. Also influencing the results is the fact that these measurements were done in isolated monocytes and luciferase experiments in another cell type, in a more controlled environment, without the complexity of all other genome variations.

4. Conclusion

We investigated the functional effect of *CD80* and *CD86* promoter polymorphisms in gene expression and we suggest that polymorphisms within the 3' end of *CD80* promoter significantly affect the activity of the promoter, decreasing the activity of the haplotype 4, likely through differential binding of transcription factors. The *CD86* SNP -298T>C did not

interfere in gene expression under the tested conditions. When the increase of protein expression on monocytes after a stimulus was analyzed, haplotype 3 had the smallest increase relative to non stimulated monocytes, while haplotype 4 had the second smallest response. For the *CD86* -1934A>G SNP, the A/A genotype showed a better response in protein expression. For the complete elucidation of the functionality of these polymorphisms it will be necessary first to test each one of the polymorphisms individually to identify which of the three candidate polymorphisms in the *CD80* promoter is causing a significant reduction in gene expression and then confirm these findings *in vivo*.

Acknowledgements

We thank the individuals that donate blood samples and made this work possible and our lab colleagues for the help with the graphics, especially Andressa Valengo, who made figure 1. Thanks to Dr. Sandro Luis Bonatto and Cladinara Roberts Sarturi for the valuable time spent in sequencing training in their lab. This work was supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Programa de Apoio a Núcleos de Excelência (PRONEX), Programa Institutos do Milênio - CNPq, Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Paraná and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Role of the funding source

The funding sources had no role in the study design, collection, analysis and interpretation of the data, in the writing or decision to submit the paper for publication.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] Chambers CA. The expanding world of co-stimulation: the two-signal model revisited. *Trends Immunol* 2001;22:217–23.
- [2] Keir ME, Sharpe AH. The B7/CD28 costimulatory family in autoimmunity. *Immunol Rev* 2005;204:128–43.

- [3] Bhatia S, Edidin M, Almo SC, Nathenson SG. B7-1 and B7-2: similar costimulatory ligands with different biochemical, oligomeric and signaling properties. *Immunol Lett* 2006;104:70–5.
- [4] Collins AV, Brodie DW, Gilbert RJC, Iaboni A, Manso-Sancho R, Walse B, et al. The interaction properties of costimulatory molecules revisited. *Immunity* 2002;17:201–10.
- [5] Sansom DM, Manzotti CM, Zheng Y. What's the difference between CD80 and CD86? *Trends Immunol* 2003;24:313–8.
- [6] Suvas S, Singh V, Sahdev S, Vohra H, Agrewala JN. Distinct role of CD80 and CD86 in the regulation of the activation of B cell and B cell lymphoma. *J Biol Chem* 2002;277:7766–75.
- [7] a Wong CK, Lunsw Ko FW, Ip WK, Hui DS, Lam CW. Increased expression of plasma and cell surface co-stimulatory molecules CTLA-4, CD28 and CD86 in adult patients with allergic asthma. *Clinical and Experimental Immunology* 2005; 141(1):122-29.
- [8] b Wong CK, Lit LC, Tam LS, Li EK, Lam CW. Aberrant production of soluble costimulatory molecules CTLA-4, CD28, CD80 and CD86 in patients with systemic lupus erythematosus. *Rheumatology* 2005; 44(8):989-94.
- [9] Ip WK, Wong CK, Leung TF, Lam CW. Plasma concentrations of soluble CTLA-4, CD28, CD80 and CD86 costimulatory molecules reflect disease severity of acute asthma in children. *Pediatric Pulmonology* 2006; 41(7):674-82.
- [10] Duin, DV, Allore HG, Mohanty S, Ginter S, Newman FK, Belshe RB, Medzhitov R, Shaw AC. Pre-vaccine Determination of the Expression of Costimulatory B7 Molecules in Activated Monocytes Predicts Influenza Vaccine Responses in Young and Older Adults. *Vaccine* 2007; 195:1590–7.
- [11] Fong TC, Wu Y, Kipps TJ. Identification of a promoter element that regulates tissue-specific expression of the human CD80 (B7.1) gene. *Journal of Immunology* 1996; 157(10):4442-50.

- [12] Lim W, Gee K, Mishra S, Kumar A. Regulation of B7.1 Costimulatory Molecule is Mediated by the IFN Regulatory Factor-7 through the Activation of JNK in Lipopolysaccharide-Stimulated Human Monocytic Cells. *Journal of Immunology* 2005; 175:5690-5700.
- [13] Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I, Schacherer F. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Research* 2000; 28(1):316-9.
- [14] Teutsch SM, Booth DR, Bennetts BH, Heard RNS, Stewart GJ. Association of common T cell activation gene polymorphisms with multiple sclerosis in Australian patients. *J Neuroimmunol* 2004;148:218–30.
- [15] Li J, Liu Z, Jiang S, Cortesini R, Lederman S, Suci-Foca N. T suppressor lymphocytes inhibit NF-kappa B-mediated transcription of CD86 gene in APC. *Journal of Immunology* 1999; 163(12):6386-92.
- [16] Lahiri DK, Nurnberger JI Jr. A rapid non-enzymatic method for the preparation of HMWDNA from blood for RFLP studies. *Nucleic Acids Res* 1991;19:5444.
- [17] Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning. A Laboratory Manual*. 2d ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.
- [18] Ewing B, Hillier L, Wendl MC, Green P. Base-Calling of Automated Sequencer Traces Using Phred. Accuracy Assessment. *Genome Research* 1998; 8(3):175-85.
- [19] Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Research* 1998; 8:195-202.
- [20] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 1994; 22:4673-80.

- [21] Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999;41:95–8.
- [22] Excoffier L and Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 2010; 10:564-7.
- [23] Matsushita M, Tsuchiya N, Oka T, Yamane A, Tokunaga K. New polymorphisms of human CD80 and CD86: lack of association with rheumatoid arthritis and systemic lupus erythematosus. *Genes Immun* 2000;1:428–34.
- [24] Beltrame MH, Pincerati MR, Dalla-Costa R, Wassem R, Köhler KF, Chautard-Freire-Maia EA, Tsuneto LT, Petzl-Erler ML. *CD80* and *CD86* polymorphisms in populations of various ancestries: 5 new *CD80* promoter alleles. *Human Immunology* 2011; 73:111-7.

Table 1. Differences among *CD80* promoter haplotypes

<i>CD80</i>	dbSNP	Hap 1	Hap 2	Hap 3	Hap 4
-1929T>C	rs16829988	T	T	T	C
-1907A>C	rs61692996	A	A	A	C
-1609A>G	rs11922270	G	A	G	G
-1559T>C	rs55916342	T	T	T	C
-1482_-1483insC	rs56052734	DEL	DEL	DEL	INS
-1478A>G	rs111793222	A	A	A	G
-1467C>T	rs112384456	C	C	C	T
-1360G>A	rs112746804	G	G	G	A or G
-1348C>T	rs72953001	C	C	C	T
-1197T>A	rs66797637	T	T	T	A
-890C>A	rs3915165	C	C	C	A or C
-820T>C	rs3915166	T	T	T	C or T
-774G>A	rs1852212	G	G	G	A or G
-757_-761insTTTT	None	DEL	DEL	DEL	INS
-687C>T	rs34394661	C	C	C	T
-558_-561insCATGA	None	DEL	DEL	DEL	INS
-454C>A	rs68180496	C	C	C	A
-387T>C	rs1880661	T	C	T	T
-232G>A	rs41271393	G	G	G	A
-79C>G	rs16829984	G	C	C	C
-7T>C	rs16829980	T	T	T	C
5C>A	rs41271391	C	C	C	A
288A>T	rs56124423	A	A	A	T

dbSNP: rs number for SNPs in the NCBI data base, Hap: *CD80* promoter haplotypes;

The four common promoter haplotypes (Matsushita et al., 2000) are shown, with the first nucleotide representing the most common. Four individuals presented other haplotypes that differed in one nucleotide from haplotype 4, included in the table as A or G, for example, when one individual carried the previous described haplotype 4 presented G instead of A in this position.

Table 2. Results of Ewens-Watterson neutrality test for *CD80* promoter alleles

Population	E-W p	n	k	Obs F	Exp F
GRC	0.527	76	4	0.57436	0.59680
GND	0.499	84	4	0.57774	0.60186
GKW	0.064	160	3	0.44322	0.72033
KRC	0.570	98	4	0.61875	0.60487
KIV	0.700	120	4	0.74021	0.61909
EU-B	0.140	142	5	0.35333	0.55704
AF-B	0.100	245	8	0.25514	0.42941
JA-B	0.000	89	4	0.25641	0.61065

E-W p : p values for Ewens-Watterson neutrality test; n: number of individuals, k: number of different alleles, GRC: Guarani M'byá, GND: Guarani Ñandeva, GKW: Guarani Kaiowá, KRC: Kaingang from Rio das Cobras, KIV: Kaingang from Ivaí, EU-B: Brazilians of predominant European ancestry, AF-B: Brazilians of predominantly mixed African and European ancestry, JA-B: Brazilians of Japanese ancestry; Obs F: Observed F values; Exp F: Expected F values

Table 3. Potential alterations of transcription factor binding sites introduced by the polymorphisms in the *CD80* and *CD86* promoters

Gene / Allele	Binding sites	Transcription factors
<i>CD80</i>		
-1929*T	TGTGGT, GGTCC, TCCA	AML1a, H4TF2, NF1
-1929*C	TGGC, GCCCA, CCCACA	NF1, AML1c e AML1a
-1907*A	GACACA	GR
-1907*C	None	None
-1609*A	GNGYGCA, GTGCA, AGTGGCT	MTF1, Sp1, CP2
-1609*G	GGTGG	CAC-binding protein
-1559*T	ATCACA	GR
-1559*C	GGACC, ACCACA	H4TF-2, AML1a
-1482*del	TGGCA	NF-1/L
-1482*C	None	None
-1478*A	TGGCA	NF-1/L
-1478*G	None	None
-1467*C	None	None
-1467*T	CACATG	muEBP-C2, twi, TFE3-S,
-1360*G	GGGCGG	Sp1, ETF, CP1
-1360*A	GGGCA	ER-alpha, LF-A1, Sp1
-1348*C	None	None
-1348*T	GATTTC	H4TF-1
-1197*T	AGTGGCT, TAAAAT	CP2, POU1F1a e F2F
-1197*A	TGGCA, GGCAAA	NF-1/L, Ap-2alphaA
-890*C	CAGGTG	c-Myc, USF2, E12, Tal-1
-890*A	TCTAAGG	E12
-820*T	GCTGT	GT-IIBa
-820*C	None	None
-774*G	TKNNGNAAK, TGTGGG	C/EBPbeta e delta, AML1a e
-774*A	ATGGGA	PTF1-beta
-757*20	None	None
-757*26	None	None
-687*C	None	None
-687*T	TTGGCT	NF-1
-557_-561*del	None	None
-557_-561*ins	TGAC, GCATGA	AP-1, TEF
-454*A	CAGAG, AGAGTGGC, AGTGGCT	GR, CBF(2), CP2
-454*C	None	None

Table 3. Potential alterations of transcription factor binding sites introduced by the polymorphisms in the *CD80* and *CD86* promoters

Gene / Allele	Biding sites	Transcription factors
<i>CD80</i>		
-387*T	CTTGTG	HES-1
-387*C	GGTCC, CTKTK	H4TF-2, TCFs e LEFs
-232*A	AGCAAAG, MAMAG	TCF-4E, TCFs e LEF
-232*G	CAAGGCTA	SF-1
-79*G	None	None
-79*C	None	None
-7*C	GCGA	E2F+p107
-7*T	TGAACT	GR
5*C	CCCTC	CTCF
5*A	None	None
287*A	CAAAGG, MAMAG, AAGGAA	TCF-1s, TCF-2alpha, LEF-1,
287*T	TGGA	NF-1
<i>CD86</i>		
-1934*A	TAAAT, AAATG	POU1F1a, YY1
-1934*G	None	None
-298*T	None	None
-298*C	CTTTATTCCA, ATTCCATT,	TFIID, POU1F1a, NF-1,

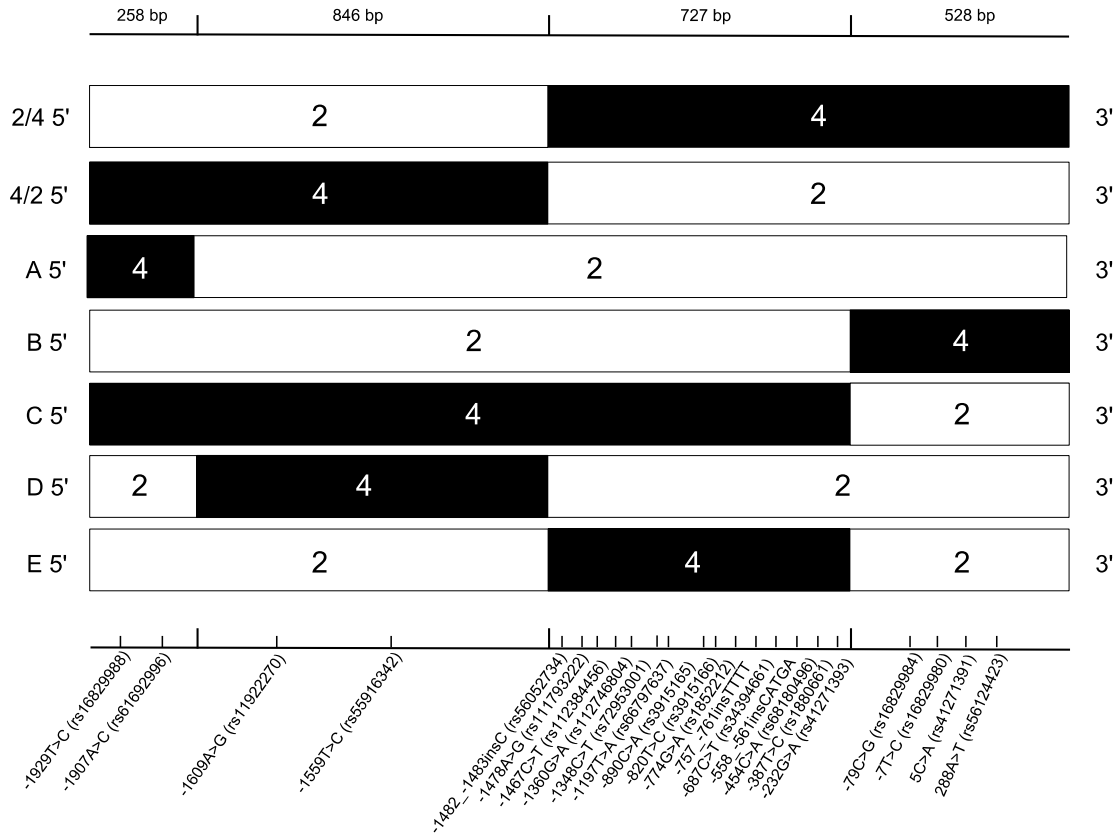


Figure 1. *CD80* recombinant haplotypes made of combinations of haplotypes 2 and 4

CD80

g.-557_-561insCATGA g.-454C>A (rs68180496) g.-387T>C
(rs1880661)

marmo GGGTGGGTGG-----CAAGCTAGCT CACTGTCCAGTGTGGCTGCCA GAGGGTGGTCTTTTGA--TGA
rhesusGGGTGGGTGG-----CAAGCTAGTT CACTGTCCAGCGTCGCTGGCA GAGGGTGATCTTGTGAGCTGA
chimp GGGTGGGTGG-----CAACCTAGTT CACTGTCCAGCGTGGCTGGCA --GGGTGGTCTTGTGAGCTGA
human GGGTGGGTGG**GCATGAC**AAACCTAGTT CACTGTCC**CAGAGTGGCT**GGCA --GGGTGGT**CC**TGTGAGCTGA

g.-232G>A (rs41271393) g.-7T>C (rs16829980)g.5C>A (rs41271391)

marmo CACAAAGCAAGGCTAGCGGGC -----
rhesus CACACAGCAAGGCTAGCGGGA TACTTCTGAGTCAATTCAAATCTTCTGTCAAG
chimp CACACAGCAAGGCTAGCCAGA ---TACTGAGCAAACCTCAAACCCTCTGTAAAG
human CACACAGCAA**AGCTA**GCCAGA ---TACTGAG**CGAACT**CAAAC**ACTCT**GTAAAG

CD86

g.-1930A>G (rs11712647) g.-298T>C (rs79111254)

marmo AAATC---AATATATATACT CCCTTATTCTATTTTGTAGA
rhesus AAATCCTAAATATATATACT ----TTATTCTATTTTGTAGC
chimp AAATCCTAAATGTATATACT CTCTTATTCTATTTTGTTC
human AAATCCT**TAAGT**GTATATACT CTCTTATT**CC**ATTTTGTTC

Figure 2. Phylogenetic conservation around some of the *CD80* and *CD86* promoter polymorphisms, as an example of how the analysis was performed.

Highlighted, the predicted TFBS; in bold and underlined, the selected polymorphisms; hyphen: deleted nucleotide; marmo: marmoset, *Callithrix jacchus*; rhesus: rhesus macaque, *Macaca mulatta*; chimp: chimpanzee, *Pan troglodytes*; human: human, *Homo sapiens*

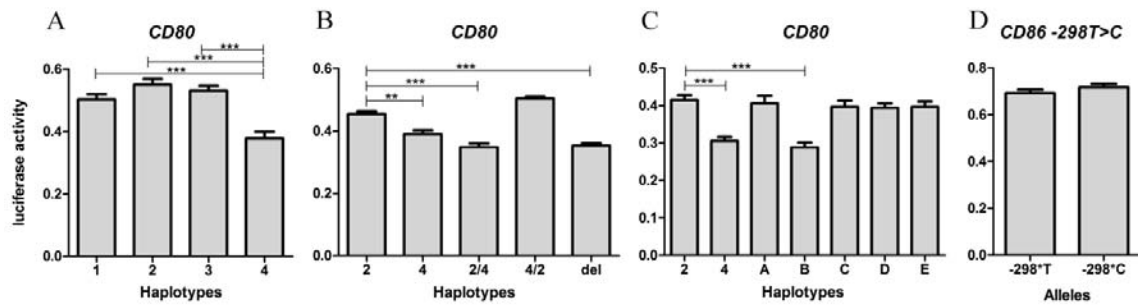
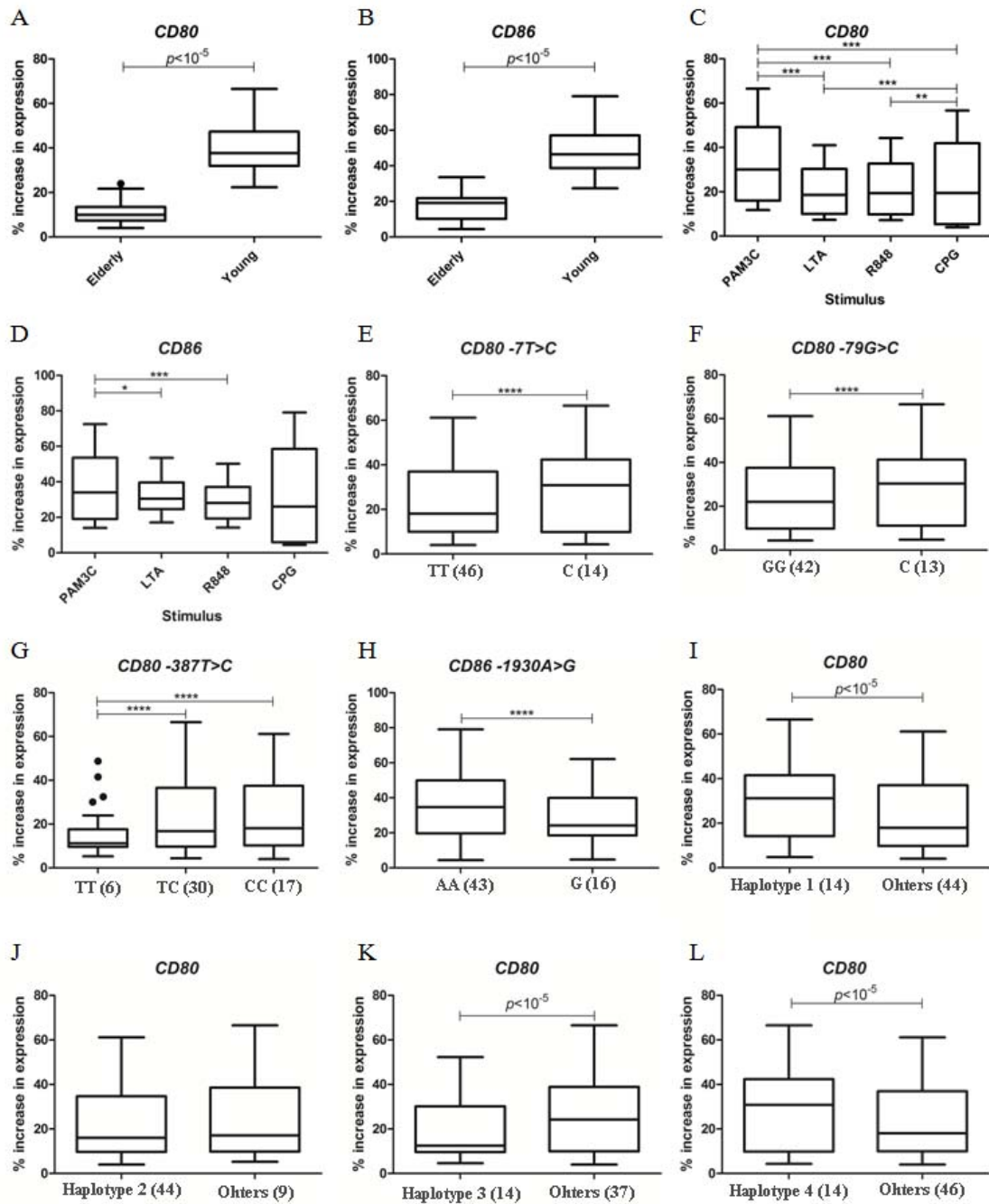


Figure 3. Luciferase measurements representing promoter activity of the *CD80* haplotypes and the *CD86* -298T>C SNP

CD80 haplotypes 1, 2, 3 and 4 are presented on table 2, haplotypes 2/4, 4/2, A, B, C, D and E are recombinant haplotypes, described on figure 1; ** $p < 0.01$, *** $p < 0.001$



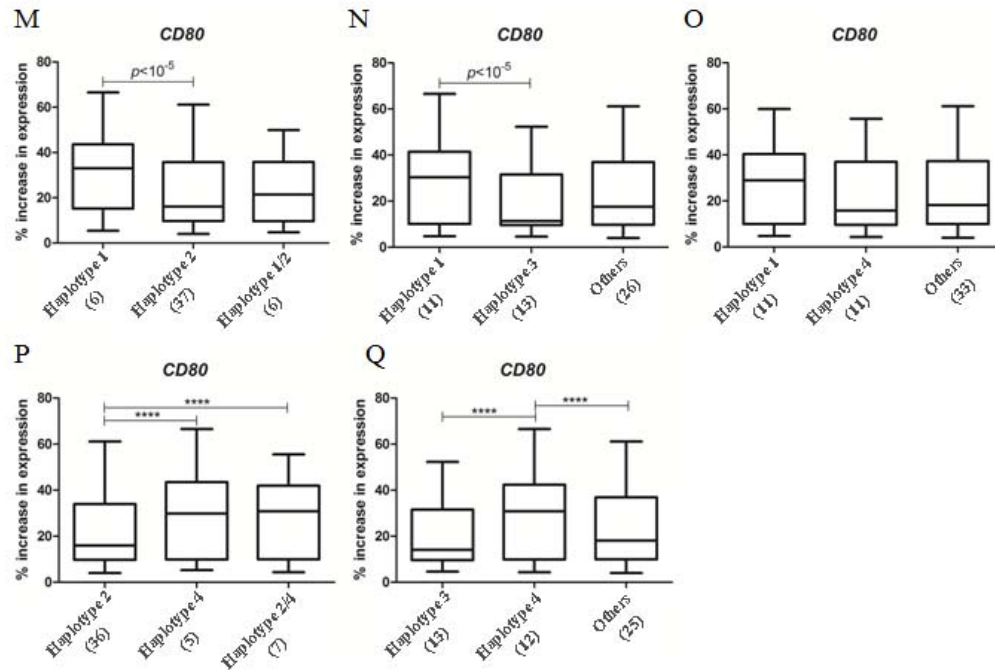


Fig. 4. Differences of CD80 and CD86 increase on the cell surface of monocytes stimulated *in vitro*, in comparison to the absence of any stimulus, for individuals with different promoter haplotypes

A and B - comparison between different age groups; C and D - differences between stimuli, PAM3C (N-palmitoylS-bis palmitoyloxy propyl cystein), LTA (lipoteichoic acid), R-848 (imidoazoquinoline) and CPG (CpG containing DNA); E to Q - comparison between genotypes considering all stimuli (some comparisons were not possible because of the small number of individuals with some haplotypes); numbers inside brackets close to the genotype and haplotype symbols represent the number of individual with that genotype or haplotype (each individual has four measurements, one for each type of stimulus) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Box plots show Tukey whiskers at 1.5 times the interquartile distance or the highest and lowest point, whichever is shorter; dots represent outliers.

5 DISCUSSÃO

No presente trabalho foram analisados polimorfismos da região promotora do gene *CD80* e exon 8 do gene *CD86* em populações de diferentes ancestralidades. As regiões promotoras e exon 1 dos genes *CD80* e *CD86* foram ainda sequenciadas em uma subamostra de indivíduos e os alelos comuns foram testado quanto ao impacto funcional na taxa de transcrição do gene.

Em um fragmento de 2,4 Kb, incluindo a região promotora e o exon 1 do gene *CD80*, foram encontrados 23 polimorfismos sendo um deles uma inserção de 4 pb ainda não descrita, que foi identificada no haplótipo 4. Essa inserção pode representar uma variação de número de cópias, VNTR (*variable number of tandem repeats*) ou microsatélite, pois um dos alelos encontrados tem uma repetição de 20 timinas e o outro, de 24 timinas, e é possível que existam ainda outros alelos com diferentes números de repetições, pois sequências repetidas tendem a mutar facilmente devido ao deslizamento da polimerase durante a replicação do DNA. Matsushita et al. (2000) descreveram quatro haplótipos na região promotora do gene *CD80*. O haplótipo 4 concentrou a grande maioria das 23 variantes encontradas, divergindo dos demais haplótipos em mais de 20 nucleotídeos, embora nem todos os cromossomos contendo o haplótipo 4 continham todas as 23 variantes descritas. É muito improvável que todas essas variações tenham surgido em um único cromossomo em regiões tão próximas, portanto é provável que haplótipos intermediários tenham existido ou ainda existam. Os quatro haplótipos já foram descritos em japoneses e australianos descendentes de europeus (MATSUSHITA et al., 2000; TEUTSCH et al., 2004), e outros 5 haplótipos foram encontrados neste trabalho em africanos, afro-brasileiros, euro-brasileiros e indígenas Kaingang (BELTRAME et al., 2011). Esses novos haplótipos correspondem, em parte, a haplótipos intermediários entre o haplótipo 4 e os demais, preenchendo parte da lacuna na história evolutiva destes polimorfismos, outros são mais provavelmente resultado de recombinação entre haplótipos já descritos. Outros haplótipos podem existir em populações ainda não analisadas, porém provavelmente os haplótipos 1, 2, 3 e 4 são os mais frequentes em todo o mundo.

Em descendentes de japoneses as frequências dos quatro haplótipos são elevadas, assim como na população japonesa (MATSUSHITA et al., 2000). AF-B também apresentaram frequências bem distribuídas, embora apresentem alguns haplótipos novos encontrados em poucos indivíduos. EU-B também possuem frequências relativamente bem distribuídas e GKW, embora não tenham o haplótipo 1, apresentam um alto nível de

heteroziguidade. O teste de Evens-Watterson revelou um valor de F observado significativamente inferior ao esperado nos JA-B, o que indica alta heteroziguidade e possível efeito de seleção balanceadora atuando na modulação da diversidade. Esse resultado pode significar, alternativamente, efeito de gargalo de garrafa recente, ou ainda, estruturação populacional. É importante notar que a população japonesa analisada por Matsushita et al. (2000) apresenta frequências muito similares aos JA-B, portanto nosso resultado não se deve a nenhum fator exclusivo de nossa população, como estrutura populacional por exemplo. As amostras de EU-B, AF-B e GWK seguiram a mesma tendência, embora as diferenças não sejam significativas. Para GWK o valor de p foi próximo ao limiar de significância. Nas demais populações indígenas o mesmo efeito não foi observado, o que pode ser devido ao grande efeito de deriva genética sofrido pelas populações ameríndias durante as migrações de entrada na América e posteriormente dentro no continente, época em que apresentaram reduzidos números efetivos populacionais.

Os haplótipos 1, 2, 3 e 4 da região promotora de *CD80* foram inseridos em vetor de expressão contendo o gene da luciferase e transfectados em células HEK293T. A expressão promovida pelo haplótipo 4 foi significativamente menor do que a dos demais. O haplótipo 4 contém a maior parte das 23 variantes menos comuns e para identificar quais causam alteração na expressão foram construídos sete haplótipos recombinantes e um mutante no qual foi deletada a inserção localizada na posição g.-557_-561insCATGA. Estes haplótipos foram construídos a partir dos haplótipos 2 e 4 e foi possível concluir que os polimorfismos com efeito na transcrição gênica localizam-se na extremidade 3' do segmento, incluindo os SNPs g.-7T>C (rs16829980), g.5C>A (rs41271391) e 287A>T (rs56124423). Um ou mais deles podem ser funcionais, mas já foi relatada a ligação do fator de transcrição AP-1 (*activating protein 1*) na região do SNP g.-7T>C (FONG, WU e KIPPS, 1996). Essa região foi determinada como uma região de ligação para fatores de transcrição através de análise com DNase I (*DNase I footprint*) e por análises *in silico* foi identificada como a proteína AP-1. Também por análises *in silico* foi determinado que o polimorfismo está localizado em uma região importante para o reconhecimento pela proteína e pode interferir na ligação (MATSUSHITA et al., 2000). Este SNP provavelmente é funcional, interferindo na expressão gênica através da eliminação de um sítio de ligação para a proteína ativadora AP-1 (o nucleotídeo só poderia ser T para que ocorresse a ligação), o que não exclui a possibilidade dos outros dois SNPs também apresentarem efeito na expressão gênica.

Os haplótipos da região promotora do gene *CD80* foram também correlacionados com variação da expressão da proteína na membrana de monócitos isolados após estímulos para

TLR. Devido ao pequeno tamanho da amostra, não foi possível comparar todos os haplótipos diretamente, mas foram encontradas diferenças significativas para todos os genótipos testados. O haplótipo 1 foi o que mostrou tendência a ter a maior resposta no aumento da transcrição, seguido pelos haplótipos 4, 3 e 2. Esses resultados são diferentes dos encontrados com os experimentos da luciferase, que representavam a taxa de expressão basal do gene, pois as células não foram estimuladas. Além disso, foram realizados em outro tipo celular (experimentos que mediram a taxa de transcrição foram feitos em linhagens celulares HEK293T, enquanto a medida das proteínas foi feita em monócitos) e células diferentes podem conter um conjunto diferente de proteínas reguladoras.

Para identificar qual ou quais dos três polimorfismos da extremidade 3' do haplótipo 4 do gene *CD80* causa(m) a redução da expressão será necessário testar cada um deles individualmente. Isso pode ser feito introduzindo uma mutação por PCR, como foi feito para testar a inserção de 5 pb, já que por recombinação não é possível, devido a falta de enzimas de restrição que cortem a sequência apenas no ponto desejado. Os alelos de cada SNP podem ainda interagir de maneira a modificar a expressão do gene, por isso seria interessante além de testar cada SNP individualmente, testar diferentes combinações entre eles. Após a identificação do polimorfismo funcional, a ligação a fatores de transcrição pode ser testada pela técnica de retardamento de migração de banda em eletroforese (EMSA, *electrophoretic mobility shift assay*), na qual sondas contendo a sequência de cada alelo são ligadas a um extrato de proteínas nucleares ou a proteínas purificadas e a mistura é submetida a eletroforese em gel de poliacrilamida, para comparação dos dois alelos quanto à migração pelo gel.

No gene *CD86* o SNP g.-298T>C não interferiu na expressão de acordo com os experimentos nos quais foi medida a taxa de transcrição, ao contrário do previsto pela análise de ligação *in silico*. Este resultado pode ser devido ao fato de que as proteínas reguladoras que se ligam a esta região não estão presentes nas células HEK293T. Também é possível que não estejam presentes em quantidades suficientes para causar diferenças mensuráveis da expressão basal do gene, sem estímulos. Já o SNP g.-1930A>G mostrou diferenças significativas de expressão de proteína na membrana de monócitos estimulados, medido por citometria de fluxo, com o genótipo A/A apresentando melhor resposta aos estímulos do que os portadores do alelo G.

6 CONCLUSÕES

As frequências dos polimorfismos analisados na região promotora de *CD80* e exon 8 de *CD86* variam significativamente entre as populações analisadas. Todos os polimorfismos investigados tiveram origem na África, antes das primeiras migrações para fora do continente. Cinco novos haplótipos ou alelos promotores de *CD80* foram descritos e são de origem provavelmente africana, europeia e ameríndia. O alelo promotor 2 é o mais provável alelo ancestral. O nucleotídeo -79 do gene *CD80* foi monomórfico em quatro das cinco populações ameríndias analisadas.

Na região promotora do gene *CD80* foram encontrados 23 polimorfismos sendo um deles uma inserção de 4 pb ainda não descrita, que foi identificada no haplótipo 4. Quatro haplótipos já tinham sido descritos em japoneses e australianos descendentes de europeus, e outros 5 haplótipos foram encontrados em africanos, afro-brasileiros, euro-brasileiros e indígenas Kaingang. Em descendentes de japoneses as frequências haplotípicas dos quatro haplótipos são similares e o teste de Evens-Watterson revelou o efeito de seleção balanceadora atuando na modulação da diversidade.

Os experimentos que mediram a taxa de transcrição revelaram uma expressão significativamente menor do haplótipo 4. O efeito é devido aos polimorfismos localizados na extremidade 3' do segmento, incluindo os SNPs g.-7T>C (rs16829980), g.5C>A (rs41271391) e 287A>T (rs56124423). Um ou mais deles podem ser funcionais, mas já foi relatada a ligação do fator de transcrição AP-1 (*activating protein 1*) na região do SNP g.-7T>C e através de análises *in silico* conclui-se que esse polimorfismo pode interferir na ligação.

As análises que correlacionaram os haplótipos de *CD80* com o aumento da expressão da proteína na membrana de monócitos isolados após estímulo por ligante de TLR mostraram diferenças significativas. O haplótipo 1 foi o que mostrou a maior resposta no aumento da expressão em relação a expressão nas células sem estímulo, seguido pelos haplótipos 4, 3 e 2.

No gene *CD86* o SNP g.-298T>C não interferiu na expressão de acordo com os experimentos que mediram a taxa de transcrição, ao contrário do previsto pela análise de ligação *in silico*. Já SNP g.-1930A>G influenciou significativamente a expressão da proteína na membrana de monócitos estimulados, com o genótipo A/A apresentando melhor resposta aos estímulos do que os portadores do alelo G.

REFERÊNCIAS

- ABDALLAH, A. M. et al. A polymorphism in the promoter region of the *CD86* (B7.2) gene is associated with systemic sclerosis. **International Journal of Immunogenetics**, Londres, v. 33, p. 155-161, 2006.
- ALEGRE, M. L.; FRAUWIRTH, K. A.; THOMPSON, C. B. T-Cell regulation by CD28 and CTLA-4. **Nature Reviews**, Londres, v. 1, p. 220-228, 2001.
- BATHIA, S. et al. B7-1 and B7-2: Similar costimulatory ligands with different biochemical, oligomeric and signaling properties. **Immunology Letters**, Amsterdam, v. 104, p. 70-75, 2006.
- BRAUN-PRADO, K. et al. HLA class I polymorphism, as characterised by PCR-SSOP, in a Brazilian exogamic population. **Tissue Antigens**, Copenhagen, v. 56, p. 417-427, 2000.
- CAVALLI-SFORZA, L. L. The DNA revolution in population genetics. **Trends in Genetics**, Cambridge, v.14, n. 2, p. 60-65, 1998.
- CAVALLI-SFOZA, L.L.; FELDMAN, M.W. The application of molecular genetic approaches to the study of human evolution. **Nature Genetics supplement**, Nova York, v. 33, p. 266-275, 2003.
- CHAMBERS, C.A.; The Expanding world of co-stimulation: the two-signal model revisited. **Trends in Immunology**, v. 22, n. 4, p. 217-223, 2001.
- COLLINS, M.; LING, V.; CARRENO, B.M. The B7 family of immune-regulatory ligands. **Genome Biology**, Londres, v. 6, p. 223.1-223.7, 2005.
- COLLINS, A.V. et al. The Interaction Properties of Costimulatory Molecules Revisited. **Immunity**, v. 17, p. 201-210, 2002.
- DALLA-COSTA, R. et al. Polymorphisms in the 2q33 and 3q21 chromosome regions including T-cell coreceptor and ligand genes may influence susceptibility to pemphigus foliaceus. **Human immunology**, v.71, n. 8, p. 809-17, 2010.
- FAUCZ FR, PROBST CM, PETZL-ERLER ML. Polymorphism of LMP2, TAP1, LMP7 and TAP2 in Brazilian Amerindians and Caucasoids: implications for the evolution of allelic and haplotypic diversity. **European Journal Immunogenetics**, v. 27, p .5-16, 2000.
- FICKETT, J.W.; WASSERMAN, W.W. Discovery and modeling of transcriptional regulatory regions. **Current Opinion in Biotechnology**, v. 11, p. 19-24, 2000.

FONG, T.C.; WU, Y.; KIPPS, T.J.; Identification of a promoter element that regulates tissue-specific expression of the human CD80 (B7.1) gene. **Journal of Immunology**, v. 157, n. 10, p. 4442-4450, 1996.

FUNDAÇÃO NACIONAL DO ÍNDIO (FUNAI). Disponível em: <<http://www.funai.gov.br>>. Acesso em: 15 jan. 2012.

HEY, J. On the Number of the New World Founders: A Population Genetic Portrait of the Peopling of the Americas. **PLOS Biology**, v.3(6) p.e193, 2005.

Instituto Brasileiro de Geografia e Estatística (IBGE). **Censo Demográfico**. Disponível em: <<http://www.ibge.gov.br/home/default.php>>. Acesso em: 15 jan. 2012.

Instituto Socioambiental. Disponível em: <<http://www.socioambiental.org/website/pib/index.html>>. Acesso em: 15 jan. 2012.

IP, W.K. et al. Plasma concentrations of soluble CTLA-4, CD28, CD80 and CD86 costimulatory molecules reflect disease severity of acute asthma in children. **Pediatric Pulmonology**, v. 41, n. 7, p. 674-82, 2006.

JEANNIN, P. et al. Soluble CD86 is a costimulatory molecule for human T lymphocytes. **Immunity**, Cambridge, v. 13, p. 303-312, 2000.

JUNE, C. H. et al. The B7 and CD28 receptor families. **Immunology Today**, Amsterdam, v. 15, p. 321-330, 1994.

KAKOULIDOU, M. et al. Human Soluble CD80 is Generated by Alternative Splicing , and Recombinant Soluble CD80 Binds to CD28 and CD152 Influencing T-cell Activation. **Scandinavian Journal of Immunology**, v. 66, p. 529–537, 2007.

LIM, W. et al. Regulation of B7.1 Costimulatory Molecule Is Mediated by the IFN Regulatory Factor-7 through the Activation of JNK in Lipopolysaccharide-Stimulated Human Monocytic Cells. **Journal of Immunology**, v. 175, p. 5690-5700, 2005.

MAGISTRELLI, G. et al. Identification of an Alternatively Spliced Variant of Human CD86 mRNA. **Biochemical and Biophysical Research communications**, Nova York, v. 280, p. 1211-1215, 2001.

MATSUSHITA, M. et al. New polymorphisms of human CD80 and CD86: lack of association with rheumatoid arthritis and systemic lupus erythematosus. **Genes and Immunity**, Houndmills, v. 7, p. 428-434, 2000.

PARHAM, P. et al. Episodic evaluation and turnover of HLA-B in the indigenous human populations of the Americas. **Tissue Antigens**, Copenhagen, v. 50, p. 219-232, 1997.

PETZL-ERLER, M.L.; LUZ, R.; SOTOMAIOR, V.S. The HLA polymorphism of two distinctive South-American Indian tribes: The Kaingang and the Guarani. **Tissue Antigens**, Copenhagen, v. 41, p. 227-237, 1993.

PROBST, C.M. et al. HLA polymorphisms and evaluation of European, African, and Amerindian contribution to the white and mulatto populations from Parana, Brazil. **Human Biology**, Detroit, v. 72, p. 597-617, 2000.

REEVES, R. H. et al. The costimulatory genes C80 and Cd86 are linked on mouse chromosome 16 and human chromosome 3. **Mammalian Genome** v. 8 p. 581-582, 1997.

SALZANO, F.M. Molecular variability in Amerindians: widespread but uneven information. **Annals of the Brazilian Academy of Sciences**, v. 74, n. 2, p. 223-263, 2002.

SANSOM, D.M.; MANZOTTI, C.M.; ZHENG, Y. What's the difference between CD80 and CD86? **TRENDS in Immunology**, Oxford, v. 24, p. 313-318, 2003.

SELVAKUMAR, A. et al. Genomic organization and chromosomal location of the human gene encoding the B-lymphocyte activation antigen B7. **Immunogenetics**, Nova York, v. 36, p. 175-181, 1992.

SUVAS, S. et al. Distinct role of CD80 and CD86 in the regulation of the activation of B cell and B cell lymphoma. **Journal of Biological Chemistry**, v. 277, n. 10, p. 7766-7775, 2002.

SCHUG, J.; OVERTON, G.C. **TESS: Transcription Element Search Software on the WWW**. Technical Report CBIL-TR-1997-1001-v0.0 Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania, 1997. URL: <http://www.cbil.upenn.edu/tess>

TEUTSCH, S. M. et al. Association of common T cell activation gene polymorphism with multiple sclerosis in Australian patients. **Journal of Neuroimmunology**, Amsterdam, v. 148, p. 218-230, 2004.

THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, p. 4673-4680, 1994.

TISHKOFF, S. et al. The genetic structure and history of Africans and African Americans. **Science**, v. 324, n. 5930, p. 1035-44, 2009.

TSUNETO, L.T. et al. HLA class II diversity in seven Amerindian populations. Clues about the origins of the Aché. **Tissue Antigens**, Copenhagen, v. 62, p. 512-526, 2003.

WANG, S. et al. Genetic variation and population structure in native Americans. **PLoS genetics**, v. 3, n. 11, p. 2049-2067, 2007.

WEG-REMERS, S. et al. Major histocompatibility complex (MHC) class III genetics in two Amerindian tribes from Southern Brazil: the Kaingang and the Guarani. **Human Genetics**, Berlin, v. 100, p. 548-556, 1997.

WINGENDER, E. et al. TRANSFAC: an integrated system for gene expression regulation. **Nucleic Acids Research**, v. 28, n. 1, p. 316-319, 2000.

a WONG, C.K. et al. Increased expression of plasma and cell surface co-stimulatory molecules CTLA-4, CD28 and CD86 in adult patients with allergic asthma. **Clinical and Experimental Immunology**, v. 141, n. 1, p. 122-129, 2005.

b WONG, C.K. et al. Aberrant production of soluble costimulatory molecules CTLA-4, CD28, CD80 and CD86 in patients with systemic lupus erythematosus. **Rheumatology**, v. 44, n. 8, p. 989-994, 2005.

ZHAO, J. et al. A cell type-specific enhancer in the human B7.1 gene regulated by NF-kappaB. **Journal of Experimental Medicine**, v. 183, n. 3, p. 777-789, 1996.

ZOU, W.; CHEN, L. Inhibitory B7-family molecules in the tumor microenvironment. **Nature Reviews Immunology**, Londres, v. 8, n. 6, p. 467-477, 2008.

APÊNDICE: DETALHAMENTO DE MATERIAL E MÉTODOS

METODOLOGIAS DO CAPÍTULO 1 – Ver página 24

SEQUENCIAMENTO

De acordo com a metodologia descrita na página 35, a região promotora dos genes *CD80* e *CD86* foi analisada através do sequenciamento direto dos produtos de PCR. Para o gene *CD80* a região amplificada foi de 2489 pb (Figura 8) e a reação de PCR foi preparada com 40 ng de DNA genômico, 0,2 µM de cada oligonucleotídeo iniciador, 1,5 U de Taq platinum DNA polimerase (Invitrogen), tampão sulfato 1X, 0,25 mM de dNTP, 1,5 mM MgCl₂, DMSO 5% em um volume final de 50 µL. O programa de PCR consistiu em um passo inicial de desnaturação a 94^oC por 1 min, 5 ciclos de desnaturação a 94^oC por 1 min, anelamento a 61,5^oC por 30 seg e extensão a 72^oC por 2 min e 30 seg, 35 ciclos de 94^oC por 1 min, 60^oC por 30 seg, 72^oC por 2 min e 30 seg e um passo final de 72^oC por 10 min.

Para o gene *CD86* a região amplificada contém 2483 pb (Figura 9) e a reação de PCR contém 40 ng de DNA genômico, 0,2 uM de cada oligonucleotídeo iniciador, 0,75 U de Taq platinum DNA polimerase (Invitrogen), tampão sulfato 1X, 0,2 mM de dNTP e 1,5 mM MgCl₂ em um volume final de 50 µL. O ciclo de PCR consiste em um passo inicial de desnaturação a 94^oC por 1 min, 5 ciclos de desnaturação a 94^oC por 1 min, anelamento a 63^oC por 30 seg e extensão a 72^oC por 2min e 30 seg, 35 ciclos de 94^oC por 1 min, 60^oC por 30 seg, 72^oC por 2 min e 30 seg e um passo final de 72^oC por 10 min.

Os oligonucleotídeos iniciadores para a reação de PCR e também para as reações de sequenciamento foram escolhidos com base na sequência dos genes disponível no banco de dados NCBI, através da análise das sequências escolhidas no programa Oligo Analyzer 3.0 (Tabela 1).


```

ggagagctgg ggtttcatca cgttggccag ctggcctcga actcctggc tcaagtgatc
cgcccacctc agcctctcaa agtgctggga ttacaggcat gagccactgt gcccagcctc
cttttaaatc ttttacttgg ttttgttttg ttttgttttg ttttccagta atataggtat
tcgggttcgg cttcctttgc ccgtcttcat ccatctcacc tccagtgtcc agaaatttac
cttagaaatc tgtggtccac aggtcatgct cctagacaca ttgtaggagg aagaggaatt
aaacacatcc cattagtaca taatcttggg agagaagcct tctgaatttt acaatcagga
ttggggaaac tctgtaaagt accttcttgg tgttcctttc ttccagcctc tttgattaag
tcacatcaaa gcgctaattg caaagttggt taatgtgtgt ttctccaagg caggttagga
tttacttggg attcttaaaa tttgaggggt tctgtgatcg tcaaaatttg tattcatgta
gtccctcatg aaatttctaa aactcggctg ggtgcagtgg ctcatgcctg taatcccagc
actttgggag gccgaggagg gtggatcaca aggtcaggag tttgagacca gctggccaa
tatggtgaaa ccccatctct actaaaaatt taaaaattag cctggcatgg tggcacgcg
ctgtagtccc agctactcag gaggtgagg cagaagaatt gcttgaacctc aggaggcggg
ggttgacagt agccaagatt gcaccactgc actccagcct gggcggcaga gaaagacttc
atctaaaaaa aaaaaaagaa aataaatttc taaaactctg tatttctttg ctcttttaag
ttgacatcta taactttcta ttactttaa tacctggaaa aaaaatcact tccagtgcac
tgtaataatg tatcaagtag aagtggctaa aatcagtcaa aacatttctt tgatctgaat
aaatggtcaa gaaaatagaa aacatgaact tacctccaat ctatgttccc ttacttgaa
aatgtaatgt tgatctacta gaggtaaggc aggaattgtc acatccatat tccttgcaac
tattcaaaat atgcatctga gaccagaaca agacacaggt gaagaagggg gctcactagg
ccccttctgc caatacagaa ggagacagtg cagtggagt ctttcatagc aagtctgccc
ttcatgattt ccaaagcttc agtattcaa atctcaggtg aataggggtg tatccactaa
gctctgccac aaacttggtg ctgagatgaa ggaagccgcc gctgtagtca gtatttctca
cccacacggg ttgctctctt tctggctggt gtgggaagga tggcgtcttt ttttttttt
tttttttgag atggagtctt gctctgttac ccaagctgga ttgtagtgtt gtgatctcgg
ctcactgcaa cccatgcaag gtgctgtcag gttccacctg tgcaaacaag tacactggtc
cagcatgggt cagctgctca ctcacagggg ctccagaatc ctacaaaggg acacacaggg
gtgggtggca acctagttgg aaatgggagt gatgggcatg actaaagtct tcctcatccc
accagagggc cactcaactg tggctagggt gaccagaatt cactgtccag cgtggctggc
aggagcccac tctctgaaga tgggtgggatt cagaggtggg agtagagttg ggtggctctg
tgagctgagc actctcccaa aagtgtcccc tacatatgag gttgacagct tattcctgca
gggactcagg agtaactctg ggtgctaaaa caaaacttga gaccagcaca gaaacagcca
gcccatgaag ctggcctccc atcacacagc aaggctagcc agatgccctt cactcttctc
gggcaccatt cttctcctcc cctaggccgc ctgggtgctg tgggtgttca tgtttgtag
tccatgcacg gtgggtgaggg aggctgctgt gtgatttgcc ccagccacag cctctacaaa
agcaactaga agaagacggc agcagagggc gcttcttcaa acaccctgtc caactcctgg
cactGACAAG TACTGAGTGA ACTCAAACCC TCTGTAAAGT AACAGAAGTT AGAAGGGGAA
ATGTCGCCTC TCTGAAGATT ACCCAAAGAA AAAGTGATTT GTCATTGCTT TATAGACTGT
AAGAAGAGAA CATCTCAGAA GTGGAGTCTT ACCCTGAAAT CAAAGGATTT AAAGAAAAG
TGGAATTTTT CTTCAGCAAg taagtacatt tactattatc ttactatggg atagcttggg
agaactagtt accagcaacc tactgcttt ttaaacaac aatacctctg ctaccacctc
ctggatcca g gcttgcctc aaaggaacc

```

Figura 5 – Sequência amplificada da região promotora e exon 1 do gene *CD80*

Fonte: GenBank

Nota: **Em vermelho**, localização dos oligonucleotídeos iniciadores da PCR (de 5' a 3').

Realçado em verde claro, localização dos oligonucleotídeos iniciadores forward.

Realçado em cinza claro, localização dos oligonucleotídeos iniciadores reversos.

Em maiúscula, exon 1.

```

gtgtctctct gggagg gagg acaggctggg gtttct catt ctgccatctt gctgacgttc
actgtggttt tgatttgcac ttccttaatg actagtgggt ttgagcacat tttcatgtga
tattggccat ttctatatct tccttgaaga aatgtctggt caactctttt gcctactttt
taattgtttg ttgtttcttc tgcaaagttg taggagtttt tttaatcata tattctggat
atcaatccct aacattttat tgttttaagg aaataaattt tgagataatt tgttatacag
caatggataa ctaataaata ctaaaaactc ctaaatgtat atactaaata tattaattag
aaaattgtaa tataaaaata ctaaaagcta aaactgtctg atattaggaa gataatttta
ctacaatata tacctttttt gtgtgtatga gacaaggctc cattatggtg cccaagatgt
tctcaaactc ctgggttcaa gcaatcttcc cacctcagcc tcttgagtag ccaggattac
aggtatgtgc cattacagcc agct ttcaca gtgtatactt gatgtactat taatgtacta
tttaatgact actaaaaagt ttatgataac aggagaaaat gcttatgcta tcttaataa
aaatttgagt gtatgcattg ataccaagaa atggaaaatt gagtgtgtac actgctacca
agaaactgaa aatgacatc ccataccttt aatatttaat atttataaca aatgagggtt
acatggattt ttttatctca tcttatttct gatgtgttat gtagttataa atactatttt
atltgcactc ctgctctgca ataattttga agatagataa cttatattat tgctgtcttc
atlttttaaa attctcaaat ccagatgaaa ttatcaaagt tgataaaaata atcattgttc
tatcttctta tttgaagatt aataattgtg ttgattttgc ttccttgctc tagttagcag
tgatgtcatc tgctatlttt aaacttctt gaaataatat atgtaatcta cttctaataa
gtltttctta ttttagcatt ttgtctaac taattataa ttatttagcc ttatttctcc
atgtttaact tgctttaaag ctgactctg gtgttttcag ccatggcttc tccattttaa
ggctatltta attcatttat tattctggaa tatatcctta aataatttat ttaggaaggc
tgtcttggtg gtggatlttc tgttgcagtt gttgtlttct tgcttcttg gtgacatatt
tctattgact tgacacttaa ctggcatctt atctaggtag ataatgctaa ttcaaaattc
tgcagatatt gttctgttgt tltttgccat ttagggttga gtaagatgcc aagttgtltt
ttgtltctct gtagtcattc tgtlttcatt ttgtltttag ctttgcttlt ggaatttaaa
atgttcaaaa tgatttgtct ggatgagaat cgatlttcat aactlttgct ttgatacact
aaacagtttg agtttctaga tgatgccat tttaatcat acgaggaaat atcttctagt
atagtttctg cttgattaat tctatgtttg tctcttaggg acatctatta atltttataat
gctgctlttt tttcagactt ctg tttcaga atattcgtt tcatcaatgt aatccttggc
tatagtagga atgaaataat aaaagcagta gcttctgtct gccctccttg gttatgcagt
ccttacagaa catctcccca tctccatcc ccccaccca gctcagtgaa actctccaca
ctttggttgt ggaaattggc agggttaggt ggctactcac tcccaatcca catccacaat
aaatcactlt ttattatctt atcaaaatct gtagaatgcc tctltattct atlttgttgc
tgcggagggt ttgtltctct ttctaattat tttatlttct aggtlttttg agggaatttc
aagaggggag atlttttatt caggctcatc ttaacgtcat gtctggaact caagctactg
aattatataat tctttaatac atatagacct acgtcaatga gtttaaaactg caaggaaagg
gttaaatltc ttcctcaagt gtggtcaaaa tctgtagaga aaagaggaa acgttctctt
aaagaaagtt agctgggtag gtatacagtc attgccgaAG AAGGCTTGCA CAGGGTGAAA
GCTTTGCTTC TCTGCTGCTG TAACAGGGAC TAGCACAGAC ACACGGATGA GTGGGGTCAT
TTCCAGATAT TAGGTCACAG CAGAAGCAGC CAAAATGGAT CCCAGTGgt gagtaataat
tcttattctt tgcagagaag ttatgagttg tgactgcagt gaaaggctga ggttgaagat
gggtgcttga tgtgtgtcct tca

```

Figura 6 – Sequência amplificada da região promotora e exon 1 do gene *CD86*

Fonte: GenBank

Nota: **Em vermelho**, localização dos oligonucleotídeos iniciadores da PCR (de 5' a 3').

Realçado em verde claro, localização dos oligonucleotídeos iniciadores forward.

Realçado em cinza claro, localização dos oligonucleotídeos iniciadores reversos.

Em maiúscula, exon 1.

Tabela 1 – Sequência dos oligonucleotídeos iniciadores para amplificação e sequenciamento dos genes *CD80* e *CD86*

Oligonucleotídeo iniciador	Denominação e sequência (de 5' a 3')
PCR <i>CD80</i>	<p>CD80PROFOR: GGAGAGCTGGGGTTTCATCACG</p> <p>CD80PROREV: GGTTCTTTGAGGGCAAGCC</p>
Reação de sequenciamento	<p>CD80PRO_1AFor: CTCAAGTGATCCGCCACCT</p> <p>CD80PRO_1ARev: CACTGCACCCAGCCGAGTTT</p> <p>CD80PRO_2BFor: ATTAGCCTGGCATGGTGGCA</p> <p>CD80PRO_2BRev: GCACCCTTCTTCACCTGTGTC</p> <p>CD80PRO_3CFor: GCCGCCGCTGTAGTCAGTAT</p> <p>CD80PRO_3CRev: GGGGACACTTTTGGGAGAGTG</p> <p>CD80PRO_4DFor: GGGACTCAGGAGTAACTCTGGG</p> <p>CD80PRO_4DRev: CAGGAGGTGGTAGCAGAGGT</p>
PCR <i>CD86</i>	<p>CD86PROFOR: GTGTCTCTCTGGGAGGGAGC</p> <p>CD86PROREV: TGAAGGACACACATCAAAGCACCAT</p>
Reação de sequenciamento	<p>CD86PRO_1AFor: GAGCACAGGCTGGGGTTTCT</p> <p>CD86PRO_1ARev: CTTGGTAGCAGTGTACACACTCAA</p> <p>CD86PRO_2BFor: CAGGTATGTGCCATTACAGCCAGCT</p> <p>CD86PRO_2BRev: CACCACCAAGACAGCCTTCC</p> <p>CD86PRO_3CFor: GCTCAGCACTGGTGTTTTCAGCCAT</p> <p>CD86PRO_3CRev: TGGGGAGATGTTCTGTAAGGACTG</p> <p>CD86PRO_4DFor: ATGCTGCCTTTTTTTCAGACTTCTG</p> <p>CD86PRO_4DRev: ACCATCTTCAACCTCAGCCTTTCAC</p> <p>CD80PRO_5EFor: AGGCTCATCTTAACGTCATGTCTGG</p>

A quantificação do produto de PCR foi feita através de eletroforese em gel de agarose 1% com 2 μ L de marcador de quantificação (*Low DNA Mass Ladder*, Invitrogen). As amostras foram misturadas a uma solução de GelRed (Biotium) 2 X em DMSO diluído em corante para corrida. O gel foi fotografado sob luz ultravioleta. Os produtos foram quantificados visualmente através da comparação com as bandas do marcador.

Após a amplificação do DNA genômico, os produtos de PCR foram purificados para eliminar os reagentes em excesso da reação de PCR e tornar viável a reação de sequenciamento. Adiciona-se 4 U da enzima Exonuclease I (EXO I, USB) e 0,7 U da enzima Fosfatase Alcalina de Camarão (SAP, USB) e ddH₂O, em um volume total de 10 μ L. A reação ocorreu a 37⁰C por 45 min. Após a purificação dos produtos de PCR, estes foram quantificados através de eletroforese em gel de agarose 1% com 2 μ L de marcador de quantificação (*Low DNA Mass Ladder*, Invitrogen), como descrito anteriormente.

A reação de sequenciamento foi realizada em placas de 96 poços, em um volume final de 10 μ L, contendo 150 ng de produto de PCR, 0,25 μ M do oligonucleotídeo iniciador e 4 μ L do MIX para sequenciamento (DYEnamic ET Terminator Cycle Sequencing Kit, GE). A reação tem um passo inicial de desnaturação de 95⁰C por 20 seg seguida por 35 ciclos de anelamento a 50⁰C por 15 seg e extensão a 60⁰C por 1 min. Após a reação de sequenciamento, uma nova purificação foi realizada antes da injeção das amostras no sequenciador. Após centrifugar a placa com as reações de sequenciamento, foram adicionados 1 μ L de acetato de amônio 7,5 M e 30 μ L de etanol 96%. A placa foi fechada com borracha própria, agitada manualmente 30 vezes e centrifugada a 2500 rcf (*relative centrifugal force*, ou força centrífuga relativa, calculada usando a velocidade e o raio da centrífuga) por 45 min a 4⁰C. A placa foi virada bruscamente para descartar o sobrenatante e centrifugada invertida, rapidamente, a 500 rcf. Adicionou-se 100 μ L de etanol 70⁰C a temperatura ambiente e é feita uma centrifugação rápida a 2500 rcf com a placa aberta. Após virar bruscamente a placa para descartar o sobrenatante, a placa é centrifugada rapidamente invertida a 750 rcf. Foram adicionados 10 μ L de solução desnaturante de carregamento (*loading solution* contendo formamida, do kit de sequenciamento da GE). A placa foi coberta com adesivo e papel alumínio, misturada por 5 min, centrifugada rapidamente a 2500 rcf e armazenada a -20⁰C. As reações de sequenciamento purificadas foram injetadas no sequenciador MegaBACE 1000 para corrida eletroforética em capilares. Os parâmetros de corrida foram 2 kV por 100 seg para injeção das amostras e 6 kV por 200 min para corrida eletroforética.

EXPRESSÃO GÊNICA UTILIZANDO GENE REPÓRTER

De acordo com a metodologia descrita nas páginas 36 a 38, as sequências dos alelos promotores de *CD80* e *CD86* foram amplificadas por PCR utilizando DNA genômico de indivíduos com cada haplótipo e oligonucleotídeos iniciadores nos quais foram adicionados sítios de clivagem para as enzimas de restrição *KpnI* e *SacI*. Os produtos de PCR foram submetidos a corrida eletroforética em gel de agarose na concentração de 1% pré-corado com brometo de etídeo 1% e então visualizados e fotografados sob luz ultravioleta para confirmar a amplificação (Figura 10). As bandas foram recortadas do gel e purificadas utilizando-se o kit Mini Elute Gel extraction (Qiagen).

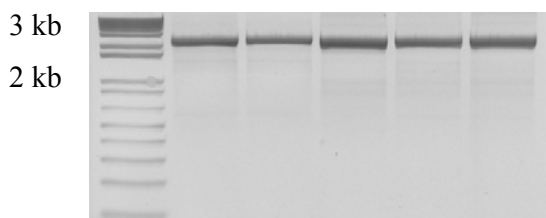


Figura 7- Gel de agarose 1% após corrida eletroforética de produtos de PCR do gene *CD80* (2359 pb) corado com brometo de etídeo e fotografado sob luz ultravioleta. Marcador de tamanho: *1 kb Plus DNA Ladder*, Invitrogen.

Os fragmentos foram então fosforilados através de uma reação contendo 10 U da enzima T4 polinucleotídeo quinase (PNK, NEB), que catalisa a transferência de gama fosfato do ATP para a extremidade hidroxil livre na extremidade 5' do DNA, tampão T4 ligase (NEB) 1X e 8 μ L do produto de PCR purificado em um volume total de 10 μ L. A reação ocorreu a 37°C por 30 min e em seguida 65°C por 20 min para inativação da enzima. Os produtos de PCR fosforilados foram purificados com o kit MinElute PCR Purification (Qiagen). A concentração dos produtos purificados foi determinada por espectrofotometria.

Os fragmentos foram ligados no vetor pSmart-HCAmp (Lucigen Corporation, Middleton, WI, USA) (Figura 11). A reação de ligação ocorreu a 16°C de um dia para o outro (em torno de 16 horas), utilizando aproximadamente 100ng do produto de PCR fosforilado e purificado, 400 U de T4 DNA ligase (NEB) e 1X do PCR pSmart HCAmp vector premix (Lucigen) em um volume final de 10 μ L.

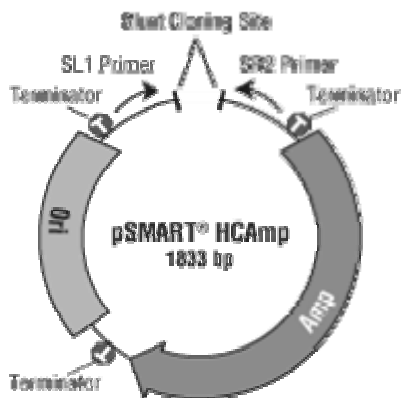


Figura 8 – Diagrama esquemático do vetor pSMART-HCAmp com *high-copy replication origins* e resistência a ampicilina

Fonte: Lucigen

Foram então adicionados 100 μ L da bactéria *E. Coli* DH5 alpha à mistura de ligação e a transfecção ocorreu por choque térmico utilizando 42^oC por 55 segundos e em seguida os tubos foram imediatamente colocados em gelo. As bactérias foram colocadas em 250 μ L do meio de cultura líquido SOC (2% triptona, 0,5% extrato de levedura, NaCl 10 mM, KCl 2,5 mM, MgCl₂ 10 mM, MgSO₄ 10 mM e glucose 20 mM), em frascos de 14 ml, próprios para cultura, e incubadas a 37^oC por no mínimo 60 minutos com agitação constante. Após a incubação as bactérias foram transferidas para placas de petri de 15 cm de diâmetro pré-aquecidas a 37^oC, contendo meio de cultura sólido (contendo triptona, extrato de levedura, glicerol, agar e ampicilina em tampão de potássio fosfato). As placas foram incubadas invertidas a 37^oC até o dia seguinte. Foram selecionadas de 6 a 8 colônias por placa, que foram transferidas para meio de cultura líquido e foram cultivadas a 37^oC sob agitação constante até o dia seguinte. O DNA plasmidial foi então extraído com o kit QIAprep Spin Miniprep (QIAGEN, Düsseldorf, Germany).

O sequenciamento para confirmação da sequência de interesse foi realizado com os oligonucleotídeos iniciadores específicos para o vetor pSMART, SL1 (5' CAGTCCAGTTACGCTGAAGTC 3') e SR2 (5' GGTCAGGTATGA TTAAATGGTCAGT 3'), que são providos como parte do kit de clonagem *CloneSmart Blunt Cloning Kit* (Lucigen Corporation, Middleton, WI, USA).

O vetor pSMART contendo o fragmento de interesse foi digerido com as enzimas de restrição *KpnI* e *SacI*, que cortam o fragmento no primer de amplificação 5' e 3' respectivamente. O vetor repórter pGL4.18 [luc2P/Neo] (Figura 12, Promega, Madison, Wisconsin, USA), que contém o gene da luciferase luc2P, do organismo *Photinus pyralis*,

O fragmento digerido do vetor pSMART e purificado foi inserido no vetor pGL4.18. A reação de ligação ocorreu a 16⁰C até o dia seguinte (aproximadamente 16 horas), utilizando 100 ng do vetor pGL4.18 defosforilado e 200 ng do inserto, 400 U de T4 DNA ligase (NEB) e 1X do tampão T4 ligase (NEB), em um volume final de 10 µL de reação. A transformação foi feita utilizando o mesmo protocolo já descrito anteriormente para inserção do vetor pSMART. Após a extração os plasmídeos foram novamente digeridos com as mesmas enzimas de restrição para confirmar o sucesso da transformação (Figura 15).

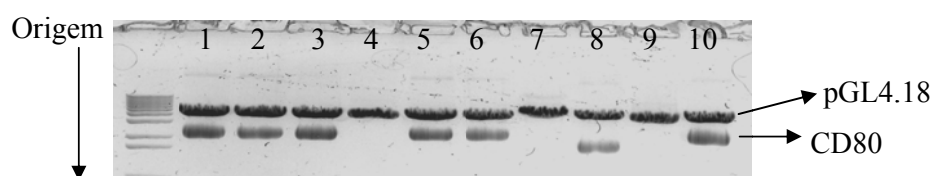


Figura 12– Gel de agarose 1% após corrida eletroforética de plasmídeos pGL4.18 (5727 pb) com inserto do gene *CD80* (2359 pb) digerido com as enzimas *KpnI* e *SacI*, corado com brometo de etídeo e fotografado sob luz ultravioleta

Colunas 1, 2, 3, 5, 6 e 10: plasmídeos com o inserto correto; 4, 7 e 9: plasmídeos sem inserto; 8: plasmídeo com inserto errado ou incompleto.

Os vetores contendo cada um dos haplótipos foram inseridos em células de uma linhagem celular humana transformada, originalmente derivada de rim de embrião (HEK293T) e temporariamente transfectado utilizando o kit Effectene transfection Reagent (Qiagen). As células foram colocadas em meio de cultura RV-2 (435 ml de DMEM, 50 ml de soro fetal bovino FBS, 10 ml de pen strep e 5 ml de L-glutamax, todos os reagentes da Invitrogen) filtrado a vácuo e foram incubadas por 48 horas a 37⁰C e 5% CO₂. A atividade da luciferase foi medida pelo sistema *Luciferase Reporter Assay* (Promega).

Para a construção do haplótipo mutante, a inserção g.-558_-561insCATGA foi deletada através de PCR (Figura 16). Foram realizadas três PCRs, duas delas utilizando DNA genômico, sendo uma com o oligonucleotídeo iniciador 5' na extremidade do fragmento e o oligonucleotídeo iniciador reverso posicionado sobre a sequência da inserção porém sem conter essa sequência, formando uma alça interna no anelamento do oligonucleotídeo à sequência alvo, e a outra PCR utilizando o oligonucleotídeo reverso na extremidade 3' e o oligonucleotídeo 5' sobre a sequência da inserção, com a mesma deletada. Na segunda PCR foram utilizados, ao invés de DNA genômico, os dois produtos de PCR das reações

a ser utilizado na montagem dos haplótipos foi recortado e purificado do gel de agarose. Para concentrar o produto, facilitando a posterior ligação, os fragmentos de gel foram combinados e purificados juntos em uma mesma reação (utilizando uma só coluna de purificação), eluídos ao final do protocolo em apenas 10 μ L e ligados novamente no vetor pGL4.18 do mesmo modo já descrito para os fragmentos inteiros. A ligação foi transformada em *E. coli*, as colônias foram triadas para confirmação de que continham o inserto completo através de digestão com as enzimas *KpnI* e *SacI* e novamente sequenciadas para confirmar que os fragmentos foram ligados corretamente.

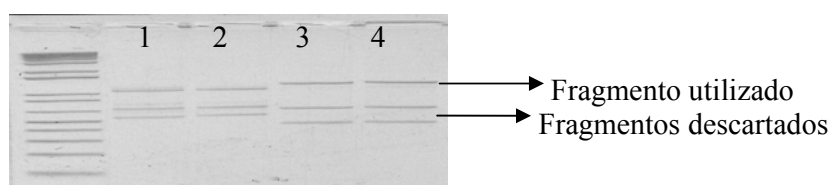


Figura 14– Gel de agarose 1% após corrida eletroforética dos haplótipos 2 e 4 do gene CD80 digeridos para a produção dos haplótipos recombinantes 2/4 e 4/2, corado com brometo de etídeo e fotografado sob luz ultravioleta

1 e 2: haplótipos 2 e 4, digeridos com as enzimas, de onde foram utilizados os fragmentos 5' (1101 pb) e descartados os demais (687 e 571 pb); 3 e 4: haplótipos 2 e 4, digeridos com as enzimas, de onde foram utilizados os fragmentos 3' (1258 pb) e descartados dos demais (662 pb e 439 pb).

ANÁLISES IN SILICO – Ver página 38

CITOMETRIA DE FLUXO – Ver página 38 e referência Duin et al., 2007.

ANEXOS

ANEXO 1.....	76
ANEXO 2.....	78

ANEXO 1

TERMO DE CONSENTIMENTO DE PARTICIPAÇÃO NO ESTUDO

“ANÁLISE DE POLIMORFISMOS GENÉTICOS, EM POPULAÇÕES”

Este é um convite para você participar voluntariamente do projeto de pesquisa ANÁLISE DE GENES DE FUNÇÃO RELACIONADA À RESPOSTA IMUNE, EM POPULAÇÕES.

Por favor, leia com atenção as informações abaixo antes de dar ou não seu consentimento para participar deste estudo. Se houver qualquer dúvida sobre o estudo ou sobre este documento, pergunte ao pesquisador com quem você está conversando neste momento.

OBJETIVO DO ESTUDO

O objetivo principal desse estudo é conhecer as variações normais de genes (formas alternativas de genes, denominadas *alelos*) e suas frequências, em populações humanas.

Por exemplo, para grupos sanguíneos, há 3 variações (alelos) principais: A, B e O. As combinações dessas resultam no grupo sanguíneo do indivíduo (combinações A/A ou A/O = grupo A; B/B ou B/O = grupo B; A/B = grupo AB; O/O = grupo O). As frequências de A, B e O diferem, entre populações. Por exemplo, nas populações indígenas das Américas o grupo sanguíneo B praticamente inexistente (quase 100% das pessoas pertencem ao tipo O), enquanto na Europa 35% das pessoas são do tipo B e, na África, 20%.

Enquanto a distribuição geográfica/étnica das variações do gene ABO são bem conhecidas, a diversidade populacional da maioria dos outros genes ainda é desconhecida. Nesse trabalho de pesquisa, vamos analisar 20 a 30 outros genes, em vários grupos da população brasileira.

O resultado será útil para conhecer as origens e as relações históricas das populações atuais e para compreender as causas da diversidade humana normal.

PROCEDIMENTOS

Se você participar deste estudo, será colhida uma amostra de seu sangue (10 até 20 ml, ou seja, correspondente a uma seringa pequena a média) e você irá responder a algumas perguntas sobre a sua origem geográfica e a origem de seus ancestrais, assim como sobre a sua saúde.

RISCOS

Não há riscos previstos nesta etapa do estudo, sendo que o único desconforto poderá ser a retirada da amostra de sangue.

BENEFÍCIOS

Não há nenhum benefício direto desta pesquisa para você, mas o conhecimento adquirido com este estudo poderá auxiliar na compreensão das causas da diversidade genética normal da espécie humana. Isso poderá contribuir para a compreensão das funções dos genes estudados, assim como para a análise das relações históricas/biológicas entre as populações humanas.

PARTICIPAÇÃO VOLUNTÁRIA

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

Embora não seja esperado, caso você tenha algum problema e não possa ir ao trabalho, você receberá um atestado para justificar a sua falta.

ESCLARECIMENTO DE DÚVIDAS

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

IDENTIFICAÇÃO

A sua identificação será mantida confidencial. Os resultados do estudo serão publicados sem revelar a sua identidade.

EQUIPE DE PESQUISADORES

Os pesquisadores envolvidos nesse projeto são:

Pesquisador responsável: Profa. Dra. Maria Luiza Petzl-Erler, UFPR

Colaboradores (2011): Profa. Dra. Karin Braun-Prado, Profa. Dra. Iíada Rainha de Souza, Profa. Dra. Danielle Malheiros Ferreira (Doutoras em Genética), Dr. Eduardo Rocha Amazonas de Almeida (Pós-Doutorando, UFPR), Márcia Holsbach Beltrame, Danilo Gardenal Augusto, Reginaldo Vieira de Sene, Liana Alves de Oliveira (Doutorandos em Genética, UFPR), Gabriel A. Cipolla, Sara Cristina Lobo Alves (Mestrandos em Genética, UFPR), Lorena Carolina Penha, Sheyla Mayumi Kuniwake (Biólogas, UFPR), Maria Áurea Santos Ribeiro (Técnica em Patologia Clínica), Eleonora Paulini, Simone Lenis Fachin, Andressa Eloisa Valengo, (Graduandas, UFPR).

Diante do exposto acima eu, _____ abaixo assinado, declaro que fui esclarecido sobre os objetivos do presente estudo, sobre os desconfortos que poderei sofrer, assim como sobre os benefícios que poderão resultar deste estudo. Concedo meu acordo de participação de livre e espontânea vontade. Foi-me assegurado o direito de abandonar o estudo a qualquer momento, se eu assim o desejar. Declaro também não possuir nenhum grau de dependência profissional ou educacional com os pesquisadores envolvidos nesse projeto (ou seja os pesquisadores desse projeto não podem me prejudicar de modo algum no trabalho ou nos estudos), não me sentindo pressionado de nenhum modo a participar dessa pesquisa.

Curitiba, ____ de _____ de _____

Voluntário (assinatura)

Documento

Pesquisador (assinatura)

RG

ANEXO 2

Laboratório de Genética Molecular Humana (LGMH)**FICHA DE AVERIGUAÇÃO****Painel do Laboratório**

CÓDIGO DO INDIVÍDUO: _____

TIPO SANGUÍNEO: _____ MATERIAL COLHIDO / VOLUME: _____

IDENTIFICAÇÃO INDIVIDUAL:

NOME: _____ IDADE: _____

TELEFONE(residencial): _____ TELEFONE(comercial): _____

TELEFONE(celular): _____ TELEFONE(fax): _____

MUNICÍPIO DE RESIDÊNCIA: _____ UF: _____

RUA: _____ Nº.: _____

BAIRRO: _____ E-MAIL: _____

MUNICÍPIO DE NASCIMENTO: _____ UF: _____

DATA DE NASCIMENTO: ____ / ____ / ____ SEXO: _____

PROFISSÃO: _____

GRAU DE INSTRUÇÃO: _____

PESSOA P/ CONTATO: _____ TELEFONE: _____

GRUPO ÉTNICO / ANCESTRALIDADE:

Laboratório	Segundo IBGE
<input type="checkbox"/> Africano (N)	
<input type="checkbox"/> Europeu (B)	Branca
<input type="checkbox"/> Indígena (I)	Preta
<input type="checkbox"/> Indígena + Branco (IB)	Amarela
<input type="checkbox"/> Mulato-claro (MC)	Parda
<input type="checkbox"/> Mulato-médio (MM)	Indígena
<input type="checkbox"/> Mulato-escuro (ME)	
<input type="checkbox"/> Oriental (O)	
<input type="checkbox"/> Oriental + Branco (OB)	
<input type="checkbox"/> Outro (descrever ao lado)	

DETALHES FAMILIARES DO INDIVÍDUO:

NOME COMPLETO DO PAI: _____

MUNICÍPIO DE NASCIMENTO DO PAI: _____ TELEFONE: _____

NOME DA MÃE (nome de solteira): _____

MUNICÍPIO DE NASCIMENTO DA MÃE: _____ TELEFONE: _____

TAMANHO DA IRMANDADE (incluindo você): ☺ TOTAL: _____ ☺ VIVOS: _____

PARA USO DO AVERIGUADOR:

ORIGEM ÉTNICA DO PAI: _____

ORIGEM ÉTNICA DA MÃE: _____

SOFRE DE ALGUMA DOENÇA? TEM IGUAL NA FAMÍLIA? (pais, irmãos, filhos): _____

ANCESTRALIDADE (descrever): _____

Laboratório	Segundo IBGE
<input type="checkbox"/> Africano (N)	
<input type="checkbox"/> Europeu (B)	Branca
<input type="checkbox"/> Indígena (I)	Preta
<input type="checkbox"/> Indígena + Branco (IB)	Amarela
<input type="checkbox"/> Mulato-claro (MC)	Parda
<input type="checkbox"/> Mulato-médio (MM)	Indígena
<input type="checkbox"/> Mulato-escuro (ME)	
<input type="checkbox"/> Oriental (O)	
<input type="checkbox"/> Oriental + Branco (OB)	
<input type="checkbox"/> Outro (descrever ao lado)	

AVERIGUADOR: _____ DATA: ____/____/____

Grupo étnico: **B:** branco com traços europeus apenas; **B br:** s / traços africanos ou indígenas claros, porém não apenas europeu; **I:** índio / indígena / ameríndio; **IB:** mestiço de índios mais branco/europeu; **MC, MM, ME:** mulato claro/ médio/ escuro (levar em conta a presença de traços europeus e também africanos); **N:** africano; **O:** oriental; **OB:** mestiço de oriental mais branco brasileiro ou europeu; se **OUTRA**, descrever.

Grau de instrução: **Ensino Fundamental** (antigo 1º grau): incompleto/ completo; **Ensino Médio** (antigo 2º grau): incompleto/completo; **Ensino Superior/Graduação:** incompleto/completo; **Pós Graduação**