

Anabelle Cavalcante Camarotti de Lima

ELABORAÇÃO DE FERRAMENTA EM BIOINFORMÁTICA PARA A
PROPOSIÇÃO DE SNPs EM GENES HUMANOS RELACIONADOS A
PATOGENIA DO HPV

AGOSTO – 2006

RECIFE – PERNAMBUCO

Anabelle Cavalcante Camarotti de Lima

ELABORAÇÃO DE FERRAMENTA EM BIOINFORMÁTICA PARA A
PROPOSIÇÃO DE SNPs EM GENES HUMANOS RELACIONADOS A
PATOGENIA DO HPV

DISSERTAÇÃO ENTREGUE AO
PROGRAMA DE PÓS-GRADUAÇÃO
EM CIÊNCIAS BIOLÓGICAS
CCB/UFPE PARA A OBTENÇÃO DO
GRAU DE MESTRE EM
BIOTECNOLOGIA.

ALUNA: ANABELLE CAVALCANTE CAMAROTTI DE LIMA

ORIENTADOR: PROF. DR. JOSÉ LUIZ DE LIMA FILHO

CO-ORIENTADOR: PROF. DR. JOÃO RICARDO MENDES DE OLIVEIRA

AGOSTO – 2006

RECIFE – PERNAMBUCO

Lima, Anabelle Cavalcante Camarotti de

Elaboração de ferramenta em bioinformática para a proposição de SNPs em genes humanos relacionados a patogenia do HPV / Anabelle Cavalcante Camarotti de Lima. – Recife: O Autor, 2006.

vi, 49 folhas : il., fig., tab.

Dissertação (mestrado) – Universidade Federal de Pernambuco. CCB. Ciências Biológicas, 2006.

Inclui bibliografia e anexo.

1. Biotecnologia 2. Bioinformática 3. Papilomavírus humano (HPV) – Patogenia 4. Polimorfismos de base única (SNPs) I. Título.

**57.08 CDU (2.ed.)
660.6 CDD (22.ed.)**

**UFPE
CCB – 2007-024**

Anabelle Cavalcante Camarotti de Lima

Banca Examinadora

Nota: Aprovada



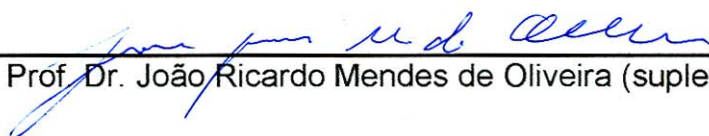
Prof. Dr. José Luis de Lima Filho (orientador)



Prof. Dr. Giovani Rota Bertani (avaliador)



Prof. Dr. Antônio Carlos Freitas (avaliador)



Prof. Dr. João Ricardo Mendes de Oliveira (suplente)

Profa. Dra. Ana Lúcia Figuerêdo Porto (suplente)

AGOSTO – 2006

RECIFE – PERNAMBUCO

“Procurar um gene dentro de um genoma humano é comparável a procurar uma pessoa sem sobrenome em uma casa sem endereço em uma rua desconhecida em uma cidade não identificada de um país estrangeiro”.

Solange Farah.

**Dedico esta dissertação as
pessoas que acreditam que
podem realizar os seus
sonhos**

AGRADECIMENTOS

Primeiramente ao meu noivo Jefferson de Barros Batista, que com todo o seu carinho e paciência, ajudou nessa empreitada.

Aos meus pais por todas as oportunidades e criação que me deram.

Ao Professor Dr. José Luiz de Lima Filho, que com sua confiança, conhecimento e ajuda psicológica auxiliou bastante no desenvolvimento desse trabalho.

Ao Professor Dr. João Ricardo Mendes de Oliveira, pela amizade e orientação a respeito das minhas idéias e atividades acadêmicas durante todo o mestrado.

Ao Professor Dr. Giovani Rota Bertani, que, mesmo tendo conhecido já no final do mestrado, ajudou imensamente na conclusão desta dissertação e na elaboração dos artigos resultantes.

A CAPES, pela concessão da Bolsa de Estudos.

Ao Professor Dr. Demetrius Araújo, do Departamento de Biologia Molecular da Universidade Federal da Paraíba, pela disponibilização do servidor que abriga o programa e pelas orientações na elaboração do artigo.

À médica. Mariléa de Lima Guimarães, pela ajuda quanto os aspectos médicos da infecção pelo HPV.

Aos Professore(a)s: Dr. Laimonis A. Laimins da Northwestern University, Feinberg School of Medicine, Chicago, USA; Dr. Aloysius J. Klingelutz da University of Iowa, USA; Dr. Derek Huntley da Division of Molecular Biosciences, Imperial College London, London, Inglaterra; Dr. Jorng-Tzong Horng da National Central University, Taiwan; Dr. Victor Gurvey da Functional Genomics Group, Utrecht, Holanda; Dra. Dorie W. Schwertz da University of Illinois, Chicago, USA; Dr. Vázquez-Ortiz Guela, National Medical Center Siglo, México; Dr. Gabor T. Marth da Boston College, USA; Dr. Anthony J. Brookes da University of Leicester, Leicester, Inglaterra; pelo envio de trabalhos que complementaram a bibliografia da dissertação e artigo e pela disponibilidade em ajudar quanto ao esclarecimento de dúvidas que surgiram no decorrer da execução deste projeto.

Em especial ao Professor Dr. Raul Cuero Prairie View A&M University, Houston, Texas, USA pela atenção, esclarecimento de dúvidas e pela concessão de valiosa bibliografia.

Às minhas antigas orientadoras Professoras Dras. Adriana Bezerra Nunes e Rosa Valéria da Silva Amorim, pelo incansável incentivo em apoiar as minhas idéias e por sempre terem a curiosidade necessária de um grande pesquisador.

Ao mestrando, do Laboratório Nacional de Computação Científica, Marx Gomes van der Linden, pelo grande suporte e ajuda computacional necessária para a conclusão das idéias que resultaram nessa dissertação.

A Aaron Rowe, mestrando da University of California , Santa Barbara, Usa, e a Jack Duvall, doutoranda da University of California, Los Angeles, USA, pelas correções da versão em inglês do artigo.

As alunas de Iniciação Científica, Nataly Amorim e Rebecca Andrade, pela grande ajuda durante o levantamento e montagem do banco de dados para a seleção dos genes que seriam utilizados para a validação do programa proposto e pelo companheirismo durante os dois anos do mestrado.

Aos pesquisadores, professores e alunos que compõem o IGEMBA2, um formidável grupo de discussão pela internet que auxilia a todos na busca por referências bibliográficas.

A todos os técnicos que respondem pelo suporte ao usuário do National Center of Biotechnology Information, pela rapidez e atenção referentes a dúvidas que surgiam sobre os dados que compõem os bancos do referido servidor.

Aos professores que compõem a banca examinadora, pela paciência e disponibilidade em ler e acrescentar observações valiosas para o enriquecimento do trabalho.

A todos os meus companheiros de turma Jaciana, Adriana, Henrique, Guilherme, Giuliana, Juliana, Morse, Daniella, Conceição, e a todos os outros que fizeram parte da primeira turma de mestrado do Programa de Pós-Graduação em Ciências Biológicas, pelos ótimos momentos que compartilhamos juntos como aulas, seminários e trabalhos em grupo.

Aos Professores que nesses dois anos de estudos de mestrado contribuíram para a minha formação, como Luiz Bezerra, Galba Takaki, Paulo

Andrade e Norma Gusmão, os quais acrescentaram muito a este trabalho na área particular de cada um. Caso tenha esquecido de algum, me perdoem.

A todos os integrantes da Pós-Graduação do CCB/UFPE, pelos momentos que passamos juntos e que contribuíram de forma direta ou indireta no transcorrer do mestrado.

Ao meu grande amigo Gladstone Almeida, pelo incentivo moral mesmo estando a quilômetros de distância.

E finalmente a todos aqueles que se não contribuíram para o desenvolvimento do trabalho, pelo menos não atrapalharam.

RESUMO

O câncer de cólon uterino é apontado mundialmente como o mais incidente em mulheres, e o HPV tem sido descrito como o principal causador dessa patogenia. Durante o processo patogênico, o HPV interfere nos mecanismos gênicos da célula, daí a grande importância dos estudos desenvolvidos com a técnica de microarranjos que têm apresentado à comunidade científica diversos genes relacionados à infecção. Como os polimorfismos de base única (SNPs) vêm sendo associados a susceptibilidade ou não a doença, o presente trabalho descreve o programa SNP8, o qual foi desenvolvido para a análise de SNPs que possam ocorrer em regiões codificadoras (CDS) do genoma humano. O programa SNP8 é de livre acesso pela internet e requer apenas informações a respeito do gene que se pretende estudar, como nome, ID e símbolos oficiais. Em seguida o programa inicia a busca e alinha as seqüências RefSeq Exons, encontradas no banco público do NCBI (<http://www.ncbi.nlm.nih.gov>) e das ESTs, encontradas no banco público do UCSC Genome Browser (genome.ucsc.edu). Os SNPs encontrados entre essas seqüências são armazenados para posteriores análises. Em um processo passo, o programa busca as seqüências de SNPs já anotadas no banco público do NCBI e as compara com os SNPs previamente armazenados pelo programa. As comparações que coincidem são descartadas e as que não coincidem são propostas como potenciais novos SNPs, os quais são alinhados com a seqüência CDS do gene em estudo para a determinação da posição desses potenciais novos SNPs no CDS. Durante a validação do programa foram encontrados muitos potenciais novos SNPs em regiões CDS, dos quais 53,12% foram não sinônimos e 46,88% foram sinônimos. Esses SNPs foram confirmados por 30 – 70% do total de ESTs analisadas para cada gene. Com a descoberta desses potenciais novos SNPs para genes importantes relacionados a patogenia do HPV, nós sugerimos uma análise de confirmação desses potenciais SNPs encontrados pelo programa SNP8, o que acarretaria numa ampliação do banco dbSNP, de onde os dados foram retirados. Devido ao programa SNP8 trabalhar com dados atualizados dos bancos de dados do NCBI e do UCSC Genome Browser, a comunidade científica está melhor informada dos diferentes SNPs que podem ser encontrados exclusivamente em regiões CDS do genoma humano.

Palavras chave: HPV, Genoma, Bioinformática, SNPs

ABSTRACT

Uterine colon cancer is globally regarded as the most common cancer type to be found in women, and HPV has been described as its main cause. During the pathogenic process, HPV interferes with the genetic mechanisms of the cell. For this reason, microarray studies are highly relevant for presenting the scientific community with numerous genes related to the infection. Single base polymorphisms (SNPs) have been associated with susceptibility, or lack thereof, to uterine colon cancer. The present work describes SNP8, a software tool developed for the analysis of SNPs that may occur in coding regions (CDS) of the human genome. SNP8 can be accessed for free by anyone through a web interface and the only input it requires from the user is some information about the gene to be studied, such as name, ID and official symbols. The program searches for SNPs by aligning the RefSeq Exons sequences found on the NCBI public database (<http://www.ncbi.nlm.nih.gov>) with the ESTs found in the UCSC Genome Browser public database (genome.ucsc.edu). SNPs found between these sequences are stored for posterior analysis. For the next step, SNP8 looks for SNP sequences previously annotated on NCBI and compares them with the ones it found, discarding from its list any matching results. The remaining ones are proposed as potential new SNPs. Those are aligned with the CDS sequence of the gene in study to determine their positions in the CDS. Many potential new SNPs in CDS regions were found during experimental runnings of the program, of which 53,12% were synonyms and 46,88% non-synonyms. These SNPs were confirmed for 40-70% of the total ESTs analyzed for each gene. With the discovery of these potential new SNPs for important genes related to HPV pathogeny, we suggest an analysis for their confirmation, which would result in the expansion of the dbSNP database, the original source for this data. Since SNP8 works with data constantly updated directly from the NCBI and UCSC Genome Browser databases, the scientific community now has a tool to constantly track SNPs that can exclusively occur in the CDS regions of the human genome.

Key words: HPV, Genome, Bioinformatic, SNPs

LISTA DE SIGLAS e ABREVIATURAS

- SPRR1A – “small proline-rich protein 1A [*Homo sapiens*]”
- DEFB1 – “defensin, beta 1 [*Homo sapiens*]”
- IFI27 – “interferon, alpha-inducible protein 27 [*Homo sapiens*]”
- KLK7 – “kallikrein 7 (chymotryptic, stratum corneum) [*Homo sapiens*]”
- SPRR2E – “small proline-rich protein 2E [*Homo sapiens*]”
- CDKN1A – “cyclin-dependent kinase inhibitor 1A (p21, Cip1) [*Homo sapiens*]”
- TRIM22 – “tripartite motif-containing 22 [*Homo sapiens*]”
- ALDH1A1 – “aldehyde dehydrogenase 1 family, member A1 [*Homo sapiens*]”
- KLK6 – “kallikrein 6 (neurosin, zyme) [*Homo sapiens*]”
- IFIT1 – “interferon-induced protein with tetratricopeptide repeats 1 [*Homo sapiens*]”
- SERPINB2 – “serpin peptidase inhibitor, clade B (ovalbumin), member 2 [*Homo sapiens*]”
- FOS – “v-fos FBJ murine osteosarcoma viral oncogene homolog [*Homo sapiens*]”
- MVD – “mevalonate (diphospho) decarboxylase [*Homo sapiens*]”
- S100A8 – “S100 calcium binding protein A8 (calgranulin A) [*Homo sapiens*]”
- PCR – Reação em Cadeia da Polimerase
- EST – Sequências Expressas únicas
- SNP – Polimorfismo de base única
- HPV – Papilomavírus Humano
- NIC – Neoplasia intraepitelial cervical

LISTA DE FIGURAS

	Pág.
Figura 1 - Esquema da disposição do genoma do HPV.....	14
Figura 2 – Demonstração esquemática da agressividade.....	16
Figura 3 – Esquema hipotético do papel do HPV.....	17
Figura 4 – Esquema da ação dos genes do HPV.....	19
Figura 5 – Demonstração esquemática do processo.....	38

LISTA DE TABELAS

Pág.

Tabela 1 – Importantes exemplos da associação entre a doença e o genótipo.....22

Tabela 2 – Lista de genes que foram descritos.....36

Tabela 3 – Caracterização dos SNP melhor descrito.....42

Tabela 4 – Relato de diferentes polimorfismos já descritos.....42

JUSTIFICATIVA

Considerando as repercussões médico-sociais da infecção pelo HPV e a alta prevalência/incidência do câncer ginecológico na nossa região, é de importância estratégica que possamos definir os painéis moleculares relacionados com esta doença, o que constituiria numa possibilidade de planejar métodos para minimizar a incidência do câncer cervical, bem como a mortalidade e morbidade das pacientes infectadas.

Esta investigação preliminar dentre outras realizadas pelo grupo visa gerar dados de pesquisa para uma posterior etapa de elaboração de vacinas terapêuticas para o HPV, de maneira similar ao que nosso grupo vem realizando para o vírus da imunodeficiência adquirida (HIV) (Lu *et al.*, 2004; de Souza *et al.*, 2006).

SUMÁRIO

RESUMO.....	i
ABSTRACT.....	ii
LISTA SIGLAS E ABREVIATURAS.....	iii
LISTA DE FIGURAS.....	iv
LISTA DE TABELAS.....	v
JUSTIFICATIVA.....	vi
APRESENTAÇÃO.....	7
CAPÍTULO I.....	8
<i>INTRODUÇÃO.....</i>	8
<i>REFERÊNCIA.....</i>	10
CAPÍTULO II.....	14
<i>REVISÃO DE LITERATURA.....</i>	14
HPV.....	14
IMPORTÂNCIA DO HPV PARA O CÂNCER.....	16
IMPORTÂNCIA DA TÉCNICA DE MICROARRANJOS PARA O ESTUDO DO HPV	20
POLIMORFISMOS GENÉTICOS HUMANOS.....	20
APLICAÇÃO DA BIOINFORMÁTICA NO ESTUDO DE POLIMORFISMOS.....	22
<i>REFERÊNCIA.....</i>	25
CAPÍTULO III.....	35
<i>MATERIAIS E MÉTODOS.....</i>	35
SELEÇÃO DE ESTs.....	35
ANÁLISE <i>IN SILICO</i>	37

CAPÍTULO IV.....	39
<i>ARTIGO.....</i>	39
CAPÍTULO V.....	40
<i>CONSIDERAÇÕES FINAIS.....</i>	40
<i>REFERÊNCIAS.....</i>	43
ANEXOS.....	46

APRESENTAÇÃO

O presente trabalho utilizou dados de expressão gênica através de microarranjos para selecionar genes relacionados à patogenicidade do HPV objetivando a identificação de potenciais novos SNPs presentes em genes humanos. Para alcançar esse objetivo foi montado o programa SNP8, o qual busca em banco de dados públicos todas as informações necessárias para a análise. Funciona de forma gratuita pela internet, não requer programas auxiliares para o seu funcionamento e apresenta uma interface simples e acessível a leigos na área de bioinformática.

Visando contemplar uma visão geral do assunto e agilizar a publicação dos resultados obtidos, esta dissertação está estruturada em uma introdução e revisão bibliográfica, um tópico descrevendo a metodologia utilizada e outro descrevendo o programa desenvolvido para ajudar na solução do problema, um tópico explicando os dados obtidos e as perspectivas de futuros trabalhos.

As citações e a apresentação das referências bibliográficas foram feitas segundo as normas da Revista Bioinformatics.

CAPÍTULO I

INTRODUÇÃO

No mundo, o câncer de cólon uterino é apontado como o segundo tipo de câncer mais incidente em mulheres (Ministério da Saúde/INCA, 2006a) e, em Pernambuco, Brasil, alcança uma estimativa de 22,16 casos para cada 100.000 mulheres para o corrente ano, pondo o estado entre os dez estados brasileiros com maior estimativa para o câncer de cólon uterino (Ministério da Saúde/INCA, 2006b). Como principal causador desta patogenia é indicado o HPV (Walboomers, 1999; Eluf Neto *et al.*, 1994), que juntamente com outros fatores, é responsável no desenvolvimento da doença (Parellada *et al.*, 1998; Zeferino *et al.*, 2002). Dentre esses fatores associados à patogenia do HPV, a literatura relata o fumo (Harris *et al.*, 2004), a idade (Sarian *et al.*, 2004) e o histórico para doenças sexualmente transmissíveis (Sellors *et al.*, 2003).

Atualmente, estudos de microarranjos também estão contribuindo para a determinação desses fatores, através da análise simultânea de milhares de genes que tiveram o número de transcritos aumentados ou diminuídos, quando em presença do genoma do HPV (Chang e Laimins, 2000; Thomas *et al.*, 2001; Ahn *et al.*, 2004) ou de plasmídeos expressando as proteínas E6 e E7 do HPV16 (Duffy *et al.*, 2003).

A análise por microarranjos dessas células infectadas pelo HPV possibilitou a descoberta de diversos genes que tiveram a sua regulação de expressão alterada, de maneira direta ou indireta, em decorrência da patogenia. Para esses

genes, tem-se analisado os SNPs, os quais se encontram em grande quantidade distribuídos pelo genoma humano (Freudenberg-Hua *et al.*, 2003; Wang *et al.*, 2005a). O seu estudo se dá por técnicas laboratoriais (Useche *et al.*, 2001) ou por técnicas em bioinformática, que podem ser desde a busca e organização da enorme quantidade de dados já existentes (Brookes *et al.*, 2000; Wang *et al.*, 2005b; Xu *et al.*, 2005; Hemminger *et al.*, 2006) até a proposição de novos SNPs a partir da análise de ESTs (Nickerson *et al.*, 1997; Marth *et al.*, 1999; Barker *et al.*, 2003; Huntley *et al.*, 2006).

A maioria das ferramentas em bioinformática úteis para essas análises de busca e organização ou proposição de novos SNPs são disponibilizadas gratuitamente para ambiente Linux, tais como HGBASE (Brookes *et al.*, 2000), SNP Hunter (Wang *et al.*, 2005b) e Sean (Huntley *et al.*, 2006); ao passo que outras, mesmo sendo escritas em Linux, podem ser utilizadas diretamente via internet, possibilitando ao usuário utilizá-las a partir de qualquer programa de busca de endereços eletrônicos independente do ambiente para onde o programa tenha sido escrito (Xu *et al.*, 2005; van der Linden *et al.* – dados não publicados).

As ESTs, utilizadas na proposição dos novos SNPs, podem ser encontradas em bancos privados (Guryev *et al.*, 2005), ou em bancos públicos, como o Golden Path Genome (genome.ucsc.edu). A importância do seu estudo se dá por essas seqüências de ESTs representarem toda uma gama de informação gênica que está sendo transcrita por uma célula, quando a mesma está submetida a um determinado fator.

REFERÊNCIAS

Ahn,W.S. (2004) Searching for pathogenic gene functions to cervical cancer. *Gynecol Oncol*, **93**, 41-48.

Barker,G. *et al.* (2003) Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. *Bioinformatics*, **19**, 421-422.

Brookes,A.J. *et al.* (2000) HGBASE: a database of SNPs and other variations in and around human genes. *Nucleic Acids Res.*, **28**, 356-360.

Chang,Y.E. e Laimins,L.A. (2000) Microarray analysis identifies interferon-inducible genes and Stat-1 as major transcriptional targets of human papillomavirus type 31. *J. Virol.*, **74**, 4174-82.

Duffy *et al.* (2003) Microarray analysis identifies differentiation-associated genes regulated by human papillomavirus type 16 E6. *Virology*, **314**, 196-205.

Eluf-Neto,J. *et al.* (1994) Human papillomavirus and invasive cervical cancer in Brazil. *Brazilian J. Cancer*, **69**,114-119.

Freudenberg-Hua,Y. *et al.* (2003) Nucleotide Variation Analysis in 65 Candidate Genes For CNS Disorders in a Representative Sample of European Population. *Genome Res.*, **13**, 2271-2276.

Guryev,V. *et al.* (2005) CASCAD: a database of annotated candidate single nucleotide polymorphisms associated with expressed sequences. *BMC Genomics*, **6**, 10-14.

Harris, T.G. *et al.* (2004) Cigarette smoking, oncogenic human papillomavirus, Ki-67 antigen, and cervical intraepithelial neoplasia. *Am. J. Epidemiol.*, **159**, 834-842.

Hemminger,B.M. *et al.* (2006) TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits. *Bioinformatic*, **22**, 626-627.

Huntley,D. *et al* (2006) SEAN: SNP prediction and display program utilizing EST sequences clusters. *Bioinformatics*, **22**, 495-496.

Marth,G.T. *et al.* (1999) A general approach to single-nucleotide polymorphisms discovery. *Nat. Genet.*, **23**, 452-456.

Ministério da Saúde. Instituto Nacional do Câncer/INCA. Síntese de Resultados e Comentários: Estimativa INCA/2006. Site: http://www.inca.gov.br/estimativa/2006/index.asp?link=conteudo_view.asp&ID=5. Acessado em: 27 de abril de 2006a.

Ministério da Saúde. Instituto Nacional do Câncer/INCA. Síntese de Resultados e Comentários: Estimativa INCA/2006. Site: <http://www.inca.gov.br/estimativa/2006/index.asp?link=mapa.asp&ID=5>. Acessado em: 27 de abril de 2006b.

Nickerson,D.A. *et al.* (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.*, **25**, 2745-2751.

Parellada, C.I. *et al.* (1998) Fatores de risco para câncer cervical e seus precursores. *Revista Brasileira de Colposcopia*, p. 5-11.

Sarian,L.O. *et al.* (2004) Factors associated with HPV persistence after treatment for high-grade cervical intra-epithelial neoplasia with large loop excision of the transformation zone (LLETZ). *J. Clin. Virol.*, **31**, 270-274.

Sellers,J.W. *et al.* (2003) Incidence, clearance and predictors of human papillomavirus infection in women. *Can. Med. Assoc. J.*, **168**, 421-5.

Thomas,J.T. *et al.* (2001) Cellular changes induced by Low-Risk human papillomavirus type 11 in Keratinocytes that stably maintain viral episomes. *J. Virol.*, **75**, 7564-7571.

Useche,F.J. *et al.* (2001) High-Throughput Identification, database storage and analysis of SNPs in EST sequences. *Genome Inform.*, **12**, 194-203.

van der Linden, M.G. *et al.* SNP8: A web tool for identification potential new SNPs for genetic studies. (não publicado)

Walboomers,J.M. *et al.* (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol*, **189**, 12-19.

Wang,L. *et al.* (2005) SNPHunter: a bioinformatic software for single nucleotide polymorphism data acquisition and management. *BMC Bioinformatic*, **6**, 60-67b.

Wang,X. *et al.* (2005) Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. *Toxicol. Appl. Pharmacol.*, **207**, 84-90a.

Xu,H. *et al.* (2005) SNPselector: a web tool for selecting SNPs for genetic association studies. *Bioinformatic*, **21**, 4181-4186.

Zeferino,L.C. *et al.* (2002) HPV e a Neoplasia do Colo do Útero. *Femina*, **30**, 471–75.

CAPÍTULO II

REVISÃO DE LITERATURA

HPV

O gênero Papilomavírus pertence a família Papillomaviridae e é representado por mais de 100 genótipos diferentes, consistindo de vírus não envelopados, com simetria icosaédrica e 72 capsômeros, DNA circular, fita dupla, com aproximadamente 8000 pb, apresentando, em sua maioria, sete ORFs que codificam as proteínas regulatórias E1-E7 e duas ORFs que codificam as proteínas estruturais L1 e L2 (Chan *et al.*, 1995; zur Hausen, 1996) (Figura 1).

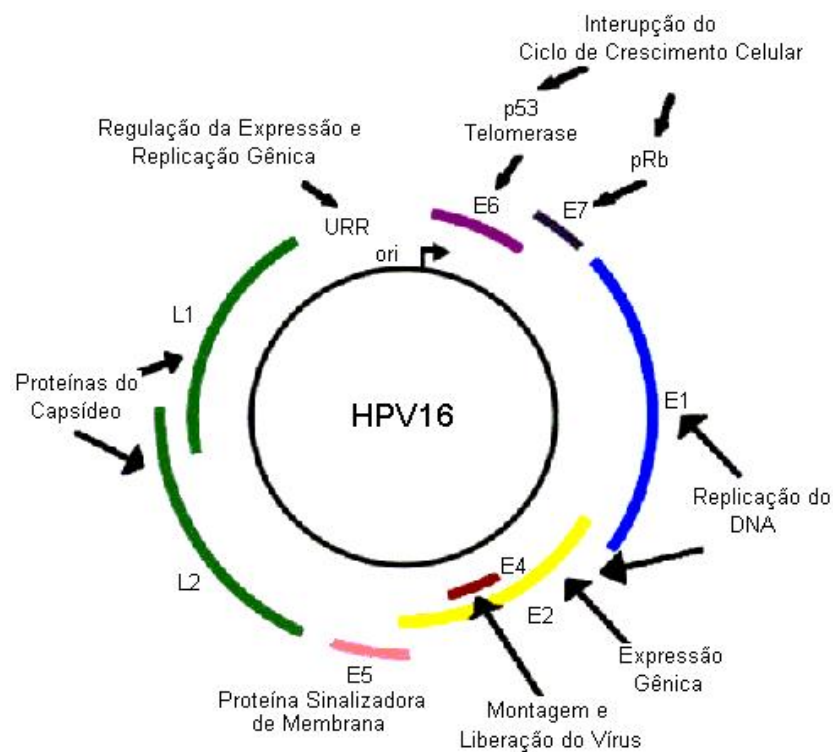


Figura 1. Esquema da disposição do genoma do HPV. São dispostas as 6 ORFs regulatórias e as 2 estruturais.

De acordo com o seu genótipo o HPV pode ser principalmente classificado em baixo, médio ou alto risco para o desenvolvimento de displasias (zur Hausen, 1996). Essa identificação ocorre através de técnicas de genotipagem que utilizam o PCR simples (Gravitt *et al.*, 1998) ou uma nova metodologia mais específica o PCR de fragmentos curtos (SF-PCR ou SPF) (Kleter *et al.*, 1998). Como exemplos de genótipos de HPV de baixo risco temos os tipos 6, 11, 41, 42, 43 e 44, considerando os de risco médio, temos os 39, 51 e 52 e os de alto risco são 16, 18, 31, 33, 35, 45 e 56 (zur Hausen, 1996; Muñoz *et al.*, 2003; de Villiers *et al.*, 2004).

A presença do HPV no hospedeiro pode ser detectada com o auxílio dos exames colposcópico (Carcopino *et al.*, 2006), citológico (Gonzalez *et al.*, 1998) e histológico (Weigl *et al.*, 2006), que são de alta sensibilidade, porém de baixa especificidade (Carvalho e Oyakawa, 2000) e muito comuns nos ambulatórios da rede pública e privada de saúde. Mais recentemente, métodos de diagnósticos mais seguros de identificação e genotipagem do HPV (Oh *et al.*, 2004) têm possibilitado, além de outras análises, a identificação do DNA-HPV no plasma sangüíneo (Pornthanakasem *et al.*, 2001), na urina de mulheres e homens (Smits *et al.*, 2005; Daponte *et al.*, 2006) e na detecção de formas latentes (Doorbar e Cubie, 2005).

Importância do HPV para o câncer

Dentre as várias possibilidades de interferência do vírus na célula hospedeira, a mais importante é o aumento da propensão ao desenvolvimento de displasia, ou desenvolvimento desordenado das células. Como a forma de transmissão mais comum do vírus é a sexual, essa displasia ocorre freqüentemente na região da cérvice feminina, sendo conhecida como neoplasia intra-epitelial cervical (NIC). (Figura 2) (Carvalho e Oyakawa, 2000).

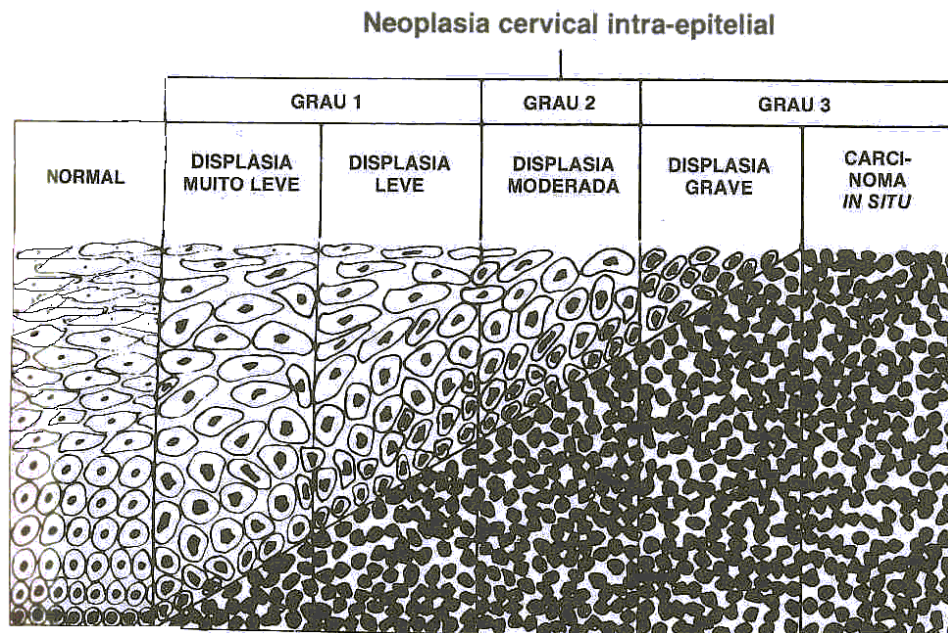


Figura 2. Demonstração esquemática da agressividade da neoplasia intra-epitelial cervical (NIC) de graus 1, 2 e 3, comparando com o tecido cervical normal (Carvalho e Oyakawa, 2000).

Quanto mais grave for o grau de displasia ou quanto maior for o potencial oncogênico do HPV, mais curto será o tempo para o desenvolvimento do carcinoma *in situ* (Cotran *et al.*, 1991). Entretanto, ressalva-se que a velocidade de progressão da doença vai variar de acordo com o genótipo do hospedeiro

(Bhattacharya e Sengupta, 2005; Trimble *et al.*, 2005), com o genótipo do vírus (Mayrand *et al.*, 2000; Gagnon *et al.*, 2004) ou com fatores ambientais (Kazimirova *et al.*, 2004; Mayrand *et al.*, 2000; Rousseau *et al.*, 2003).

Outro fator que parece interferir na progressão da doença é a forma como o DNA viral se encontra, epissomal (circular) ou integrado ao DNA do hospedeiro (Figura 3). Essa interação está relacionada a regiões descondensadas do genoma humano, ou seja, a regiões de alta frequência de transcrição, as quais são mais acessíveis a integração do DNA estranho (Klimov *et al.*, 2002).

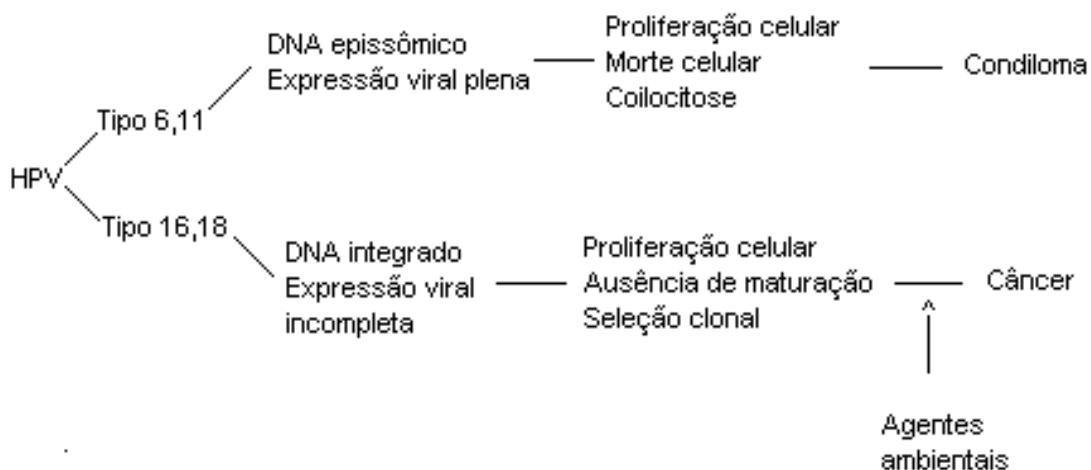


Figura 3. Esquema hipotético do papel do HPV na neoplasia do colo do útero (adaptado de Cotran *et al.*, 1991).

Durante a progressão da doença, parecem estar envolvidas as proteínas E1 e E2, que, após a integração do DNA viral ao do hospedeiro, se separam permitindo o aumento das proteínas virais E6 e E7, as quais levam à transformação neoplásica (Carvalho e Oyakawa, 2000). Outra proteína viral que vem sendo muito descrita por também estar relacionada com a transformação neoplásica é a proteína E5 (Conrad *et al.*, 1993), a qual interfere no processo de

comunicação célula-célula e no processo de comunicação interna da célula do hospedeiro (Straight *et al.*, 1993; Crusius *et al.*, 1997; Oelze *et al.*, 1995).

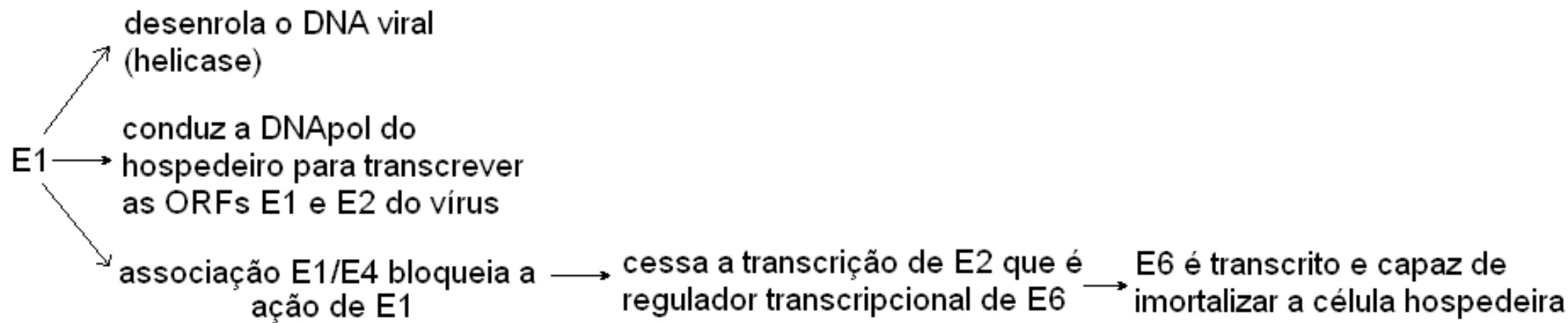


Figura 4. Esquema da ação dos genes do HPV após infecção de célula hospedeira (Interpretado a partir de Oelze *et al.*, 1995; Ashrafi *et al.*, 2006; Adam *et al.*, 2000; Disbrow *et al.*, 2003).

Importância da Técnica de Microarranjos para o Estudo do HPV

A partir de 1990, com o advento da técnica de microarranjos ou chip de DNA que detecta níveis de expressão gênica pela hibridização do mRNA com porções gênicas imobilizadas numa superfície de vidro foi possível a análise simultânea de milhares de genes que estivessem sendo expressos por uma célula, sendo útil como uma poderosa ferramenta no estudo biológico de processos internos à mesma (Schena *et al.*, 1995; Watson *et al.*, 1998).

A importância dessa técnica para a detecção da expressão gênica em células infectadas pelo HPV, além de outras vantagens possibilitou o conhecimento dos diferentes genes virais que têm maior importância para a transformação celular do hospedeiro (Duffy *et al.*, 2003; Chang e Laimins, 2000) e a descoberta de diversos genes humanos que têm a sua expressão alterada quando a célula se encontra infectada pelo genoma do HPV (Chang e Laimins, 2000; Ahn *et al.*, 2004). E então, com base nesses estudos, vem sendo possível a realização de estudos mais direcionados a determinados genes confirmadamente associados à progressão da patogenia (Lehr e Brown, 2003; Lehr *et al.*, 2004; Buck *et al.*, 2006).

Polimorfismos Genéticos Humanos

Os polimorfismos de base única (SNPs) representam aproximadamente 90% das variações gênicas comumente encontradas (Collins *et al.*, 1998; Smigielski *et al.*, 2000; Kruglyak, 2005), sendo estimada a sua ocorrência a cada

500-1000 pares de base no genoma humano (Sherry *et al.*, 1999). Esses SNPs podem ser encontrados em seqüências de promotores (Menin *et al.*, 2006), introns (Fujita *et al.*, 2006), exons (Soriano *et al.*, 2005) ou em qualquer outra região do genoma (Hudson, 2003; Wang *et al.*, 2005a).

Um dos possíveis resultados do polimorfismo é a alteração na estrutura alélica do mRNA (Shen *et al.*, 1999), podendo gerar alteração na estrutura da proteína codificada. Portanto, como a interação proteína hospedeiro/proteína viral é um fator importante durante o processo infeccioso ocasionado pelo HPV (Chang e Laimins, 2000), é de grande importância o estudo dessas mutações para a melhor compreensão dessas interações.

Os SNPs têm sido úteis na construção de mapas gênicos (Carlson *et al.*, 2001) e vêm sendo associados à diversidade populacional (Shah *et al.*, 2006), respostas individuais a tratamentos terapêuticos (Kurland *et al.*, 2005), individualidade ou exposição ao ambiente (Wang *et al.*, 2005a) e na identificação de muitas doenças onde haja interferência direta do genótipo (Tabela 1).

Tabela 1. Importantes exemplos da associação entre a doença e o genótipo do indivíduo.

Gene^a	SNP	Região Gênica	Associação	Referência
F5	G/A	Intron	Trombose venosa	Bertina <i>et al.</i> , 1994.
CCND1	A/G	Intron	Susceptibilidade a carcinogênese na infecção pelo HPV	Catarino <i>et al.</i> , 2005.
FAS	A/G	Promotor	Susceptibilidade a carcinogênese na infecção pelo HPV	Ueda <i>et al.</i> , 2005.
FAS	G/A	Promotor	Susceptibilidade a carcinogênese na infecção pelo HPV	Lai <i>et al.</i> , 2005.
GSBS	T/C	Promotor	Hipercolesterolemia	Ono <i>et al.</i> , 2003.
CCR5	T/C	Promotor	Transmissão perinatal do HIV-1	Kostrikis <i>et al.</i> , 1999
CCR5	A/G	Promotor	Transmissão perinatal do HIV-1	de Souza <i>et al.</i> , 2006
UGT1A9	T/C	Intron	Câncer de ovário e colonretal	Fujita <i>et al.</i> , 2006

^aNome Oficial baseado no Comitê de Nomenclatura Gênica (HUGO): F5 - coagulation factor V (proaccelerin, labile factor) [*Homo sapiens*]; CCND1 - cyclin D1 [*Homo sapiens*]; FAS – FAS (TNF receptor superfamily, member 6) [*Homo sapiens*]; GSBS - chromosome 7 open reading frame 16 [*Homo sapiens*]; CCR5 - chemokine (C-C motif) receptor 5 [*Homo sapiens*]; UGT1A9 - UDP glucuronosyltransferase 1 family, polypeptide A9 [*Homo sapiens*].

Com o intuito de auxiliar no desenvolvimento das pesquisas envolvendo SNPs relacionados a doenças humanas, respostas a medicamentos e a fatores ambientais foi criado, em outubro de 2002, o Consórcio Internacional HapMap, que reúne cientistas e agências financiadoras do Canadá, Japão, Reino Unido, Estados Unidos, Nigéria e China. Abrangendo diferentes etnias, pretende-se identificar as diferentes regiões cromossômicas compartilhadas pelas variações gênicas (The International HapMap Consortium, 2005).

Aplicação da Bioinformática no Estudo de Polimorfismos

A literatura nos revela que existem cerca de 9 milhões de SNPs já identificados no genoma humano e prevê ainda a busca de outras seqüências a serem identificadas (Freudenberg-Hua *et al.*, 2003; Wang *et al.*, 2005a). Até o

presente momento, esses diferentes SNPs já identificados têm sido armazenados em vários bancos de dados internacionais, como dbSNP (Sherry *et al.*, 1999), JSNP (Hirakawa *et al.*, 2001), SNP consortium (<http://snp.cshl.org>) e HapMap project (<http://www.hapmap.org>). Já a busca por novos SNPs tem ocorrido por dois métodos, o experimental, o qual consiste de um elevado número de passos laboratoriais, tornando-o dispendioso e complexo (Useche *et al.*, 2001); e o computacional, através do uso de ferramentas em bioinformática, o qual tem sido proposto por muitos pesquisadores.

A maioria das ferramentas hoje em dia disponíveis para o estudo de SNPs tem como objetivo a busca e organização dos dados já existentes, como exemplo: HGBASE (Brookes *et al.*, 2000), SNP Hunter (Wang *et al.*, 2005b), SNPselector (Xu *et al.*, 2005) e Tamal (Hemminger *et al.*, 2006), ao passo que outras têm sugerido a proposição e avaliação de novos SNPs diretamente a partir de ESTs (Nickerson *et al.*, 1997; Marth *et al.*, 1999; Barker *et al.*, 2003; Huntley *et al.*, 2006).

As ESTs são seqüências únicas formadas por 200 – 300 pb que representam um gene específico, que foi expresso pela célula em resposta a algum fator. Essas seqüências têm sido utilizadas pelo Projeto Genoma como marcadores para a construção de um mapa físico do genoma de determinada célula (Roberts e Lifton, 2004), como também vêm sendo úteis na detecção de SNPs relacionados a diferentes estados celulares, como infecção viral (Catarino *et al.*, 2005; Lai *et al.*, 2005), patologias autoimunes (Amoli *et al.*, 2005) e doenças psiquiátricas (Oliveira *et al.*, 2000; Nishimura *et al.*, 2004; Nishimura *et al.*, 2005).

A proposição de novos SNPs a partir de ESTs com a utilização de ferramentas em bioinformática tem se dado ou pela análise de banco de ESTs próprios, como o faz o programa CASCAD, que propõe SNPs apenas nos organismos *Rattus norvegicus* (rato de laboratório) e *Danio rerio* (peixe zebra) os quais tiveram os seus bancos de dados de ESTs montados pelo mesmo grupo (Guryev *et al.*, 2005); ou pela análise de banco de ESTs públicos, como o banco do UCSC Genome Browser (genome.ucsc.edu) utilizado pelo programa SNP8, o qual propõe SNPs em genes humanos através do alinhamento das ESTs com as seqüências referência de Exons (Pruitt *et al.*, 2003) obtidas do banco de dados do NCBI (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene>) (van der Linden *et al.*, 2006 – dados não publicados).

REFERÊNCIAS

Adam,J.L. *et al.* (2000) A mutagenic analysis of the E5 protein of human papillomavirus type 16 reveals that E5 binding to the vacuolar H⁺-ATPase is not sufficient for biological activity, using mammalian and yeast expression systems. *Virology*, **272**, 315-325.

Ahn,W.S. (2004) Searching for pathogenic gene functions to cervical cancer. *Gynecol Oncol*, **93**, 41-48.

Amoli,M.M. *et al.* (2005) MCP-1 gene haplotype association in biopsy proven giant cell arteritis. *J. Rheumatol.*, **32**, 507-510.

Ashrafi,G.H. *et al* (2006) Down-regulation of MHC class I is a property common to papillomavirus E5 proteins. *Virus Res.*, **120**, 208-211.

Barker,G. *et al.* (2003) Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. *Bioinformatics*, **19**, 421-422.

Bertina,R.M. *et al.* (1994) Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature*, **369**, 64-67.

Bhattacharya,P. e Sengupta,S. (2005) Lack of evidence that proline homozygosity at codon 72 of p53 and rare arginine allele at codon 31 of p21, jointly mediate cervical cancer susceptibility among Indian women. *Gynecol. Oncol.*, **99**, 176-182.

Brookes,A.J. *et al.* (2000) HGBASE: a database of SNPs and other variations in and around human genes. *Nucleic Acids Res.*, **28**, 356-360.

Buck,C.B. *et al.* (2006) Human alpha-defensins block papillomavirus infection. *Proc Natl Acad Sci U S A*, **103**, 1516-1521.

Carcopino,X. *et al.* (2006) Determination of HPV type 16 and 18 viral load in cervical smears of women referred to colposcopy. *J. Med. Virol.*, **78**, 1131-1140.

Carlson,C.S. *et al.* (2001) SNPping on the human genome. *Curr. Opin. Chem. Biol.*, **5**, 78-85.

Carvalho,J.J.M. e Oyakawa,N. (2000) *I Consenso Brasileiro de HPV: Papilomavírus humano*. BG Cultural, São Paulo, 150 p.

Catarino,R. *et al.* (2005) Increased risk of cervical cancer associated with cyclin D1 gene A870G polymorphism. *Cancer Genet. Cytogenet.*, **160**, 49-54.

Chan,S.Y. *et al.* (1995) Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny and taxonomy. *J. Virol.*, **69**, 3074-3083.

Chang,Y.E. e Laimins,L.A. (2000) Microarray analysis identifies interferon-inducible genes and Stat-1 as major transcriptional targets of human papillomavirus type 31. *J. Virol.*, **74**, 4174-82.

Collins,F.S. *et al.* (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res*, **8**, 1229-1231.

Conrad,M. *et al.* (1993) The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein. *J. Virol.*, **67**, 6170-6178.

Cotran,R.S. *et al.* (1991) *Robbins: Patologia Estrutural e Funcional*. Guanabara Koogan, Rio de Janeiro, 1231p.

Crusius,K. *et al.* (1997) Enhancement of EGF- and PMA-mediated MAP kinase activation in cells expressing the human papillomavirus type 16 E5 protein. *Oncogene*, **15**, 1437-1444.

Daponte,A. *et al.* (2006) Evaluation of high-risk human papillomavirus types PCR detection in paired urine and cervical samples of women with abnormal cytology. *J. Clin. Virol.*, **36**, 189-193.

de Souza, P.R. *et al.* (2006) CCR5 promoter polymorphisms and HIV-1 perinatal transmission in Brazilian children. *J. Reprod. Immunol.*, **69**, 77-84.

de Villiers, E-M. *et al.* (2004) Classification of papillomaviruses. *Virology*, **324**, 17-27.

Disbrow,G.L. *et al.* (2003) Codon optimization of the HPV-16 E5 gene enhances protein expression. *Virology*, **311**, 105-114.

Doorbar,J. e Cubie,H. (2005) Molecular basis for advances in cervical screening. *Mol. Diagn.*, **9**, 129-142.

Duffy *et al.* (2003) Microarray analysis identifies differentiation-associated genes regulates by human papillomavirus type 16 E6. *Virology*, **314**, 196-205.

Freudenberg-Hua,Y. *et al.* (2003) Nucleotide Variation Analysis in 65 Candidate Genes For CNS Disorders in a Representative Sample of European Population. *Genome Res.*, **13**, 2271-2276.

Fujita,K. *et al.* (2006) Novel Single Nucleotide Polymorphism of UGT1A9 Gene in Japanese. *Drug Metab. Pharmacokinet.*, **21**, 79-81.

Gagnon,S. *et al.* (2004) Viral polymorphism in human papillomavirus types 33 and 35 and persistent and transient infection in the genital tract of women. *J. Infection Disease*, **190**, 1575-1585.

Gonzalez,S.J.L. *et al.* (1998) Cytologic correlation between the Bethesda system and colposcopic biopsy. *Gynecol. Obstet. Mex.*, **66**, 330-334.

Gravitt,P.E. *et al.* (1998) Genotyping of 27 human papillomavirus types by using L1 consensus PCR products by a single-hybridization, reverse line blot detection method. *J. Clin. Microbiol.*, **36**, 1998.

Guryev,V. *et al.* (2005) CASCAD: a database of annotated candidate single nucleotide polymorphisms associated with expressed sequences. *BMC Genomics*, **6**, 10-14.

Hemminger,B.M. *et al.* (2006) TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits. *Bioinformatic*, **22**, 626-627.

Hirakawa,M. *et al.* (2002) JSNP: a database of common gene variations in the Japanese population. *Nucleic Acid Res.*, **30**, 158-162.

Hudson,T.J. (2003) Wanted regulatory SNPs. *Nature Genet.*, **33**, 439-440.

Huntley,D. *et al* (2006) SEAN: SNP prediction and display program utilizing EST sequences clusters. *Bioinformatics*, **22**, 495-496.

Kazimirova,A. *et al.* (2004) M. Does a vegetarian diet influence genomic stability? *Eur. J. Nutri.*, **43**, 32-8.

Kleter,B. *et al.* (1998) Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. *Am. J. Pathol.*, **153**, 1731-1739.

Klimov, E. *et al.* (2002) Human papillomavirus and cervical tumours: mapping of integration sites and analysis of adjacent cellular sequences. *BMC Cancer*, **2**, 24-32.

Kostrikis,L.G. *et al.* (1999) A polymorphism in the regulatory region of the CC-chemokine receptor 5 gene influences perinatal transmission of human immunodeficiency virus type 1 to African–American infants. *J Virol*, **73**, 10264–10271.

Kruglyak,L (2005) Power tools for human genetics. *Nature Genet*, **37**, 1299-1300.

Kurland,L. *et al.* (2005) Hypertension and SNP genotyping in antihypertensive treatment. *Cardiovasc. Toxicol.*, **5**, 133-142.

Lai,H-C *et al.* (2005) Genetic polymorphisms of FAS and FASL (CD95/CD95L) gene in cervical carcinogenesis: an analysis of haplotype and gene-gene interaction. *Gynecol. Oncol.*, **99**, 113-118.

Lehr,E. e Brown,D.B. (2003) Infection with the oncogenic human papillomavirus type 59 alters protein components of the cornified cell envelope. *Virology*, **309**, 53-60.

Lehr,E. *et al.* (2004) Infection with Human Papillomavirus alters expression of the small proline rich proteins 2 and 3. *J. Med. Virol.*, **72**, 478-483.

Marth,G.T. *et al.* (1999) A general approach to single-nucleotide polymorphisms discovery. *Nat. Genet.*, **23**, 452-456.

Mayrand,M.H. *et al.* (2000) Detection of human papillomavirus type 16 DNA in consecutive genital samples does not always represent persistent infection as determined by molecular variant analysis.*J. Clin. Microbiol.*, **38**, 3388-93.

Menin,C. *et al.* (2006) Association between MDM2-SNP309 and age at colorectal cancer diagnosis according to p53 mutation status. *J. Natl Cancer Inst.*, **98**, 285-288.

Muñoz,N. *et al.* (2003) Epidemiologic Classification of Human Papillomavirus Types Associated with Cervical Cancer. *N. Engl. J. Med.*, **348**, 518-527.

Nickerson,D.A. *et al.* (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.*, **25**, 2745-2751.

Nishimura,A.L. *et al.* (2005) Monoamine Oxidase A Polymorphisms in Brazilian Patients: Risk factor for Late-onset Alzheimer 's Disease?. *J. Molecular Neuroscience.*, **27**, 13-218.

Nishimura,A.L. *et al.* (2004) Lack of Association Between the Brain-Derived Neurotrophin Factor (C-270T) Polymorphism and Late-Onset Alzheimer's Disease (LOAD) in Brazilian Patients. *J Molecular Neuroscience.*, **22**, 257-260.

Oelze,I *et al.* (1995) Human Papillomavirus Type 16 E5 Protein Affects Cell-Cell Communication in an Epithelial Cell Line. *J. Virol.*, **69**, 4489-4494.

Oh,Y.K. *et al.* (2004) Enhanced mucosal and systemic immunogenicity of human papillomavirus-like particles encapsidating interleukin-2 gene adjuvant. *Virology*, **328**, 266-273.

Oliveira,J.R.M. *et al.* (2000) Analysis of the functional serotonin transporter polymorphism (5-HTTLPR) in Brazilian patients affected by dysthymia, major depression and bipolar disorder. *Molecular Psychiatry*, **5**, 348-349.

Ono,S. *et al.* (2003) A promoter SNP (-1323T>C) in 6-substrate gene (GSBS) correlated with hypercholesterolemia. *J. Human Genet.*, **48**, 447-450.

Pornthanakasem,W. *et al.* (2001) Human papillomavirus DNA in plasma of patients with cervical cancer. *BMC Cancer*, **1**. (revista em formato eletrônico)

Pruitt,K.D. *et al.* (2003) NCBI Reference Sequence project: update and current status. *Nucleic Acid Res.*, **31**, 34-37.

Roberts,R. e Lifton (2004) The Human Genome and its Future Implications for Cardiovascular Disease. *In* Fuster,V. *et al.* (eds), *Hurt's: The Heart*. IRL Press, The McGraw-Hill Companies, Inc.

Rousseau,M-C. *et al.* (2003) Predictors of cervical co-infection with multiple human papillomavirus types. *Cancer Epidemiol. Biomark. Prev.*, **12**, 1029-1037.

Schena,M. *et al.* (1995) Quantitative monitoring of genes expression patterns with a complementary DNA microarray. *Science*, **270**, 467-470.

Shah,P.R. *et al.* (2006) No association of DYNC1H1 with sporadic ALS in a case-control study of a northern European derived population: a tagging SNP approach. *Amyotroph. Lateral Scler. Other Motor Neuron Disord.*, **7**, 46-56.

Shen,L.X. *et al.* (1999) Single nucleotide polymorphisms can cause different structural folds of mRNA. *Proc. Natl. Acad. Sci.*, **96**, 7871-7876.

Sherry,S.T. *et al.* (1999) dbSNP – Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res.*, **9**, 677-679.

Smigielski,E.M. *et al.* (2000) dbSNP: a database of single nucleotide polymorphisms. *Nucleic. Acids Res.*, **28**, 352-355.

Smits,P.H. *et al.* (2005) High prevalence of human papillomavirus infections in urine samples from human immunodeficiency virus-infected men. *J. Clin. Microbiol.*, **43**, 5936-5939.

Soriano,A. *et al.* (2005) Polymorphisms in the interleukin-4 receptor alpha chain gene influence susceptibility to HIV-1 infection and its progression to AIDS. *Immunogenetics*, **57**, 644-654.

Straight,S.W. *et al.* (1993) Enhancement of EGF- and PMA-mediated MAP kinase activation in cells expressing the human papillomavirus type 16 E5 protein. *J. Virol.*, **67**, 4521-4532.

The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature*, **437**, 1299-1320.

Thomas,J.T. *et al.* (2001) Cellular changes induced by Low-Risk human papillomavirus type 11 in Keratinocytes that stably maintain viral episomes. *J. Virol.*, **75**, 7564-7571.

Trimble,C.L. *et al.* (2005) Spontaneous regression of high-grade cervical dysplasia: effects of human papillomavirus type and HLA phenotype. *Clin. Cancer Res.*, **11**, 4717-4723.

Ueda,M. *et al.* (2005) FA gene promoter –670 polymorphism (A/G) is associated with cervical carcinogenesis. *Gynecol. Oncol.*, **98**, 129-133.

Useche,F.J. *et al.* (2001) High-Throughput Identification, database storage and analysis of SNPs in EST sequences. *Genome Inform.*, **12**, 194-203.

van der Linden, M.G. *et al.* SNP8: A web tool for identification potential new SNPs for genetic studies. (não publicado)

Wang,L. *et al.* (2005) SNPHunter: a bioinformatic software for single nucleotide polymorphism data acquisition and management. *BMC Bioinformatic*, **6**, 60-67b.

Wang,X. *et al.* (2005) Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. *Toxicol. Appl. Pharmacol.*, **207**, 84-90a.

Watson,A. *et al.* (1998) Technology for microarray analysis of gene expression. *Curr. Opin. Biotechnol.*, **9**, 609-614.

Weigl,G. *et al.* (2006) Investigation of 208 consecutive cases of cervical cone biopsies with regard to indication, negative samples and quality control. *Acta Cytol.*, **50**, 185-190.

Xu,H. *et al.* (2005) SNPselector: a web tool for selecting SNPs for genetic association studies. *Bioinformatic*, **21**, 4181-4186.

zur Hausen,H. (1996) Papillomavirus infections – a major cause of human cancers. *Biomed. Biochim. Acta*,**1288**, 55-78.

CAPÍTULO III

MATERIAIS E MÉTODOS

Seleção de Genes

A etapa inicial do trabalho foi revisar principais trabalhos relacionados a infecção por HPV envolvendo estudos de expressão gênica em grande escala utilizando microarranjos de publicações mais relevantes no assunto. Dessa forma foi possível identificar diversos genes que tiveram a sua expressão desregulada em presença ou do genoma do HPV (Chang e Laimins, 2000; Thomas *et al.*, 2001; Ahn *et al.*, 2004) ou de plasmídeos expressando as proteínas E6 e E7 do HPV16 em cultivo celular (Duffy *et al.*, 2003). Dentre esses trabalhos, alguns genes foram encontrados repetidos, contudo com a quantificação da expressão diferenciada (Tabela 2), estando essas diferenças relacionadas ao genótipo do HPV utilizado no estudo. Os genes listados na Tabela 2 foram o ponto de partida para as análises *in silico*.

Tabela 2 Lista de genes que foram descritos em trabalhos de microarranjos, como relacionados à infecção pelo HPV e selecionados para as análises *in silico*.

Gene [#] (ID)	Quantificação da expressão diferencial do gene ^b			
	Localização	Duffy <i>et al.</i> , 2003 ^c	Chang e Laimins, 2000 ^c	Ahn <i>et al.</i> , 2004 ^c Thomas <i>et al.</i> , 2001 ^c
SPRR1A (6698)	chr1:151223181-151224913	-8.5 ^d	-5,4	
DEFB1 (1672)	chr8:6715511-6722939		-3,2	2.4 ^e
IFI27 (3429)	chr14:93646832-93652786		-10,1 ^e	4.4
KLK7 (5650)	Chr19: 56171541-56178962	-50	-3,8	
SPRR2E (6704)	chr1:151332236-151333625	-4.9	-5,1	
CDKN1A (1026)	chr6:36754465-36763086	-2 ^d	-2,6 ^e	
TRIM22 (10346)	chr11:5667664-5688668	-2.1	-2,8	
ALDH1A1 (216)	Chr9: 74705407-74757789	6		-4,1
KLK6 (5653)	chr19:56153700-56164741	-7.9	-2.3	
IFIT1 (3434)	chr10:91142358-91153582		-9,6 ^e	3.1
SERPINB2 (5055)	Chr18: 59705922-59722100		-3,4	2.5
FOS (2353)	chr14:74815284-74818663		-2,4	-3,5
MVD (4597)	Chr16: 87245849-87256996		-2,3	-2,2
S100A8 (6279)	chr1:151629133-151630173			76.6 ^e 2.5

^aO símbolo do gene está de acordo com o Comitê de Nomenclatura Gênica (HUGO) e a localização está de acordo com o NCBI Map Viewer.

^bA expressão diferencial de cada gene foi encontrada pela média da razão dos níveis de expressão das amostras versus os controles negativos (Chang e Laimins, 2000; Thomas *et al.*, 2001; Duffy *et al.*, 2003; Ahn *et al.*, 2004).

^cReferência onde foi analisado o gene pela técnica de microarranjos.

[#]SPRR1A – “small proline-rich protein 1A [Homo sapiens]”; DEFB1 – “defensin, beta 1 [Homo sapiens]”; IFI27 – “interferon, alpha-inducible protein 27 [Homo sapiens]”; KLK7 – “kallikrein 7 (chymotryptic, stratum corneum) [Homo sapiens]”; SPRR2E – “small proline-rich protein 2E [Homo sapiens]”; CDKN1A – “cyclin-dependent kinase inhibitor 1A (p21, Cip1) [Homo sapiens]”; TRIM22 – “tripartite motif-containing 22 [Homo sapiens]”; ALDH1A1 – “aldehyde dehydrogenase 1 family, member A1 [Homo sapiens]”; KLK6 – “kallikrein 6 (neurosin, zyme) [Homo sapiens]”; IFIT1 – “interferon-induced protein with tetratricopeptide repeats 1 [Homo sapiens]”; SERPINB2 – “serpin peptidase inhibitor, clade B (ovalbumin), member 2 [Homo sapiens]”; FOS – “v-fos FBJ murine osteosarcoma viral oncogene homolog [Homo sapiens]”; MVD – “mevalonate (diphospho) decarboxylase [Homo sapiens]”; S100A8 – “S100 calcium binding protein A8 (calgranulin A) [Homo sapiens]”.

Análises *in silico*

As análises *in silico* referem-se as etapas de bioinformática utilizando diferentes bases de dados e o programa SNP8 desenvolvido em colaboração com a Universidade Federal da Paraíba. Essas bases de dados abrangem os banco de dados públicos NCBI (<http://www.ncbi.nlm.nih.gov>) e o UCSC Genome Browser (genome.ucsc.edu), usados como fonte de informação para a proposição dos novos SNPs. Esses bancos foram escolhidos por serem descritos como abrangentes e atualizados periodicamente.

A figura 5 apresenta esquematicamente o processo realizado pelo programa SNP8 para a proposição dos novos SNPs. Resumidamente, as atividades de bioinformática inicialmente requereram poucas informações a respeito do gene que se pretende analisar, as quais podem ser obtidas no Comitê de Nomenclatura Gênica (HUGO - <http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl>). Em seguida o programa iniciou a busca das RefSeq Exons no banco público do NCBI e das RefSeq ESTs no banco público do UCSC Genome Browser, as quais foram posteriormente alinhadas e os SNPs encontrados entre elas armazenados. Em seguida, foram procuradas as seqüências de SNPs já anotadas no NCBI, as quais foram comparadas com os SNPs previamente armazenados pelo programa. As comparações que coincidiram foram descartadas e as que não coincidiram foram propostas como possíveis novos SNPs, que foram alinhados com a RefSeq CDS do gene para a determinação da posição dos possíveis novos SNPs no CDS.

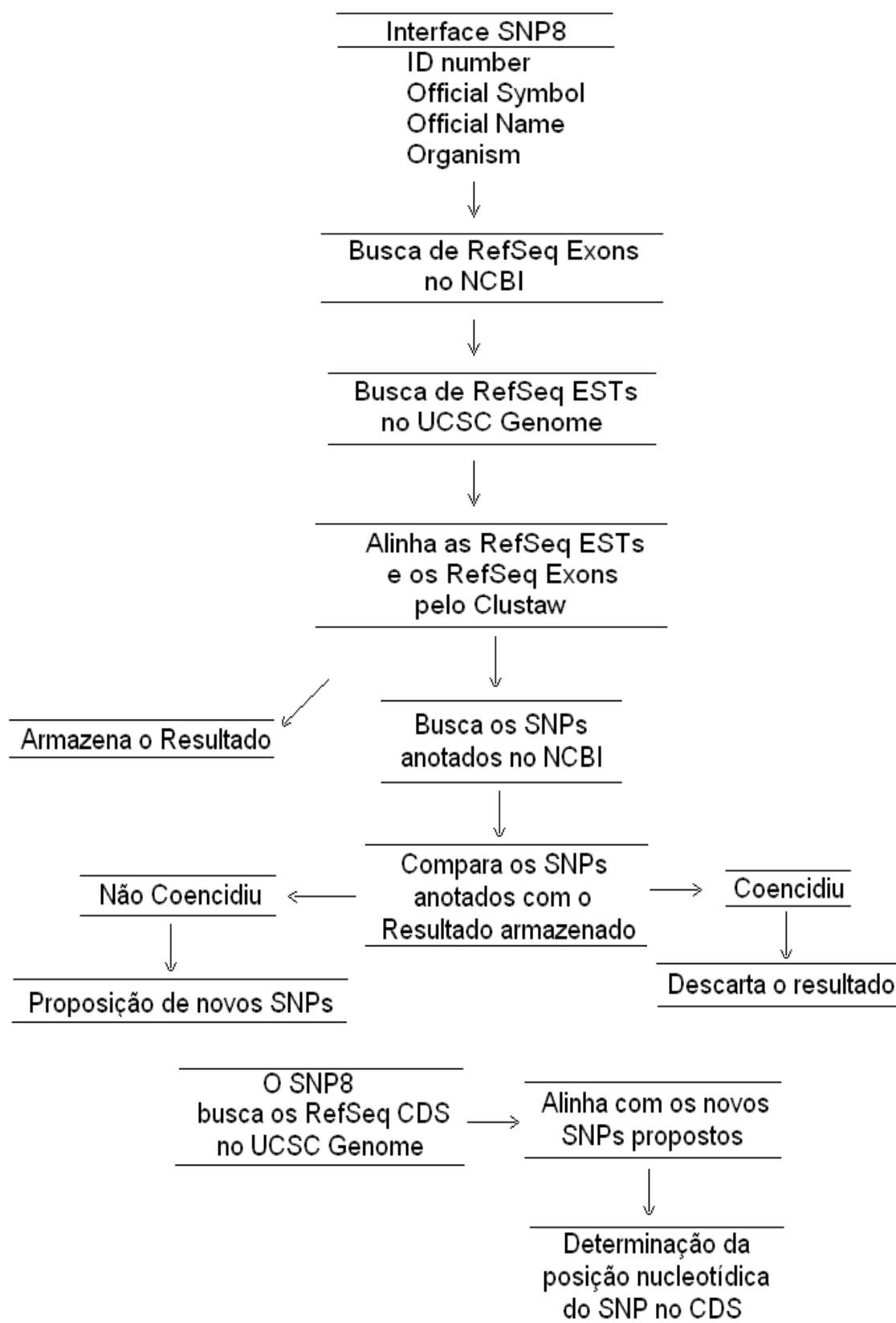


Figura 5. Demonstração esquemática do processo realizado pelo programa SNP8.

CAPÍTULO IV

ARTIGO

SNP8: A web tool for identification potential new SNPs for genetic studies

Van der Linden, M. G.¹, Camarotti-Lima, A. C.^{3*}, Araújo, D. A. M.¹,
Bertani, G. R.^{2,3}, Oliveira, J. R. M.^{3,4} and Lima Filho, J. L.^{2,3}

¹ Department of Molecular Biology, Bioinformatic Laboratory, Federal University of Paraíba, João Pessoa – Brazil, ² Department of Biochemistry, Federal University of Pernambuco (UFPE), Recife, Brasil, ³ Keiso Asami Immunopatology Laboratory, UFPE, Recife – Brazil, ⁴ Department of Neuropsychiatry, UFPE, Recife – Brazil

Trabalho submetido à revista

Bioinformatic 2006.

Sequence Analysis

SNP8: A web tool for identification potential new SNPs for genetic studies

Van der Linden, M. G.¹, Camarotti-Lima, A. C.^{3*}, Araújo, D. A. M.¹, Bertani, G. R.^{2,3}, Oliveira, J. R. M.^{3,4} and Lima Filho, J. L.^{2,3}

¹ Department of Molecular Biology, Bioinformatic Laboratory, Federal University of Paraíba, João Pessoa – Brazil, ² Department of Biochemistry, Federal University of Pernambuco (UFPE), Recife, Brazil, ³ Keiso Asami Immunopathology Laboratory, UFPE, Recife – Brazil, ⁴ Department of Neuropsychiatry, UFPE, Recife – Brazil

ABSTRACT

Motivation: SNP8 is a software tool designed for identification of Single Nucleotide Polymorphisms (SNPs) that might occur at the coding sequencing (CDS). It requires as inputs only the Official Gene name, ID, and the name of the studied organism, and can be freely accessed from the web without any need to download files or install programs. The differences between the EST and RefSeq Exon sequences of each gene are aligned with the sequences already submitted to the dbSNP to identify potential new SNPs, which have never been submitted before on dbSNP. Isoforms can also be analyzed. A final data sheet is generated in a table format for proper analysis of the relevant findings.

Availability: SNP8 is available free at <http://www.dbm.ufpb.br/bioinfo/snp8/>
Contact: bellecamarotti@yahoo.com.br
Supplementary information: <http://www.dbm.ufpb.br/bioinfo/snp8/index.php?p=supdata>

1 INTRODUCTION

Single Nucleotide Polymorphisms (SNPs) are the most common forms of genetic variation. Several internet databases search for them systematically in order to understand complex and polygenic disorders such as autoimmune syndromes, cancer and body reaction towards infectious agents (Hughes et al. 2006; Vang et al. 2005; Smigielski et al. 2000).

SNPs can be found in exons (Soriano et al. 2005), promoters (Menin et al. 2006), introns (Fujita et al. 2006) or in any other kind of sequence in eukaryotic cells (Wang et al., 2005b).

At the moment, different SNPs have been reported and compiled at various international databases such as dbSNP (<http://www.ncbi.nlm.nih.gov/SNP>), JSNP (<http://snp.ims.u-tokyo.ac.jp>), SNP consortium (<http://snp.cshl.org>) and HapMap project (<http://www.hapmap.org>). After the dramatic increase of SNPs reported from a diversity of organisms, there is a growing incentive to develop new bioinformatics tools to help organizing and analyzing the plethora of data already available at public and private databases.

The rising number of tools available to study SNPs aims at organizing and allowing data mining for all SNPs that have been already predicted and confirmed, such as HGBASE (Brookes et al., 2000), SNP Hunter (Wang et al., 2005a), SNP selector (Xu et al., 2005) and Tamal (Hemminger et al., 2006). However, is still important to build new bioinformatics tools to identify and validate potentially new SNPs (Nickerson et al. 1997, Marth et al. 1999, Barker et al. 2003, Huntley et al. 2006).

In silico SNPs detection is an inexpensive and efficient way to propose SNPs from expressed sequence tags (ESTs) of different organisms. Most current software tools for SNP studies, however, require local installation and complex requirements for data input and configuration (Useche et al. 2001, Guryev et al. 2005, Huntley et al. 2006).

In order to address some of these issues, SNP8 was developed as a simple and easy tool to propose potential new SNPs from public genomic databases.

2 METHODS

SNP8 is a web based Bioinformatics tool written in PHP and Asynchronous Javascript (AJAX) that may be used through an web interface to analyze SNPs that occur in CDS sequences, requiring of each gene only the Official Gene name and ID and the name of the studied organism based on HUGO nomenclature (<http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl>). SNP8 partial results are shown live during the execution.

To propose potential new SNPs the program searches for ESTs that have been spliced from UCSC Genome Browser (genome.ucsc.edu), whose identity levels range from 97% to 100% with mRNA/Genomic alignment. SNP8 then aligns each EST with each RefSeq exon, from Gen Bank (<http://www.ncbi.nlm.nih.gov>). Instances in which the EST and exon are found to differ by exactly one base mutation are considered to be SNPs.

The differences found between these two sequences are aligned with SNPs annotated to dbSNP on NCBI (Sherry et al., 1999) to determine potential new SNPs. For each potential new SNP, the following properties are displayed on a table: nucleotide mutation; codon position on CDS gene; mutation type (synonymous/nonsynonymous); and protein residue. If the gene has isoforms, all of them will be analyzed using the same parameters.

SNPs which have been mistyped as “N” at the electropherogram are ignored, because of the high probability that they occurred due to a sequencing error or calling mistake.

Using the data sheets generated by SNP8, the authors suggest the enlargement of currently public database SNPs, in which most of the annotated SNPs are in intronic or untranslated regions, assisting the scientific community with a good tool to study genetic differences found between ESTs and RefSeq exons in different situations.

In a follow-up to this work, the authors intend to update SNP8 to embrace other organisms, for which these Exon, ESTs and SNPs have been made available by public databases.

3 VALIDATION

Considering our interest in the identification of SNPs in human genes related to viral infections (Arraes, et al. 2004, Teles et al. 2005, de Souza et al. 2006, Segat et al. 2006), the *in silico* validation of SNP8 has been carried out with genes up or down regulated in cells infected with human papillomavirus. These genes have been consistently involved with HPV infection in previous mi-

*To whom correspondence should be addressed.

croarray studies (Chang and Laimins 2000, Thomas et al. 2001, Schwarze et al. 2002, Duffy et al. 2003, Ahn et al. 2004) (Tab 1).

ACKNOWLEDGEMENTS

We would like to thank CAPES and CNPq for financial support to develop this project and we are thankful to Jackie Duvall (UCLA-USA), and Aaron Rowe (UCSB-USA) for manuscript review, and Nataly Amorim (UFPE-BR) and Rebecca Andrade (UFPE-BR) for help with select genes have been consistently involved with HPV infection.

REFERENCES

- Ahn, W.S. et al. (2004) Searching for pathogenic gene functions to cervical cancer. *Gynecol. Oncol.*, **93**, 41-48.
- Arraes, L.C. et al. (2004) MBL2 mother's genotype influences HIV-1 vertical transmissions in Brazilian children. *J. Brasileiro AIDS*, **5**, 193-232.
- Barker, G. et al. (2003) Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. *Bioinformatics*, **19**, 421-422.
- Brookes, A.J. et al. (2000) HGBASE: a database of SNPs and other variations in and around human genes. *Nucleic Acids Res.*, **28**, 356-360.
- Chang, Y.E. and Laimins, L.A. (2000) Microarray analysis identifies interferon-inducible genes and Stat-1 as major transcriptional targets of human papillomavirus type 31. *J. Virol.*, **74**, 4174-4182.
- de Souza, P.R.E. et al. (2006) CCR5 promoter polymorphisms and HIV-1 perinatal transmission in Brazilian children. *J. Reprod Immunol*, **69**, 77-84.
- Duffy, C.L. et al. (2003) Microarray analysis identifies differentiation-associated genes regulated by human papillomavirus type 16 E6. *Virology*, **314**, 196-205.
- Fujita, K. et al. (2006) Novel Single Nucleotide Polymorphism of UGT1A9 Gene in Japanese. *Drug Metab Pharmacokinet*, **21**, 79-81.
- Guryev, V. et al. (2005) CASCAD: a database of annotated candidate single nucleotide polymorphisms associated with expressed sequences. *BMC Genomics*, **6**, 10-14.
- Hemminger, B.M. et al. (2006) TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits. *Bioinformatic*, **22**, 626-627.
- Hughes, L.B. et al. (2006) Racial/Ethnic Differences in Allele Frequencies of Single Nucleotide Polymorphisms in the Methylenetetrahydrofolate Reductase Gene and Their Influence on Response to Methotrexate in Rheumatoid Arthritis. *Ann Rheum Dis.*, **65**, [Epub ahead of print].
- Huntley, D. et al. (2006) SEAN: SNP prediction and display program utilizing EST sequences clusters. *Bioinformatics*, **22**, 495-496.
- Marth, G.T. et al. (1999) A general approach to single-nucleotide polymorphisms discovery. *Nat. Genet.*, **23**, 452-456.
- Menin, C. et al. (2006) Association between MDM2-SNP309 and age at colorectal cancer diagnosis according to p53 mutation status. *J. Natl Cancer Inst.*, **98**, 285-288.
- Nickerson, D.A. et al. (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.*, **25**, 2745-2751.
- Schwarze, S.R. et al. (2002) Novel pathways associated with Bypassing cellular senescence in human prostate epithelial cells. *J. Biol. Chem.*, **277**, 14877-14883.
- Segat, L. et al. (2006). IL-18 gene promoter polymorphism is involved in HIV-1 infection in a Brazilian pediatric population. *Immunogenetics*. In Press.
- Sherry, S.T. et al. (1999) dbSNP – Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res.*, **9**, 677-679.
- Smigielski, E.M. et al. (2000) dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res.*, **28**, 352-355.
- Soriano, A. et al. (2005) Polymorphisms in the interleukin-4 receptor alpha chain gene influence susceptibility to HIV-1 infection and its progression to AIDS. *Immunogenetics*, **57**, 644-654.
- Teles, F.R. et al. (2005) Trends in dengue diagnosis. *Rev. Med. Virol.*, **15**, 287-302.
- Thomas, J.T. et al. (2001) Cellular changes induced by Low-Risk human papillomavirus type 11 in Keratinocytes that stably maintain viral episomes. *J. Virol.*, **75**, 7564-7571.
- Useche, F.J. et al. (2001) High-Throughput identification, Database Storage and Analysis of SNPs in EST sequence. *Genome Informatics*, **12**, 194-203.
- Vang, T. et al. (2005) Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet.*, **37**, 1317-1319.
- Wang, L. et al. (2005) SNP Hunter: a bioinformatic software for single nucleotide polymorphism data acquisition and management. *BMC Bioinformatic*, **6**, 60-67a.

Wang, X. et al. (2005) Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. *Toxicol Appl. Pharmacol.*, **207**, 84-90b.

Xu, H. et al. (2005) SNPselector: a web tool for selecting SNPs for genetic association studies. *Bioinformatic*, **21**, 4181-4186.

Tabela 1 Quantification of potential new SNPs in human genes by SNP8 and quantification of annotated SNPs on dbSNP.

Gene (ID)	SNP8		dbSNP [*]	
	CDS		Exon	
	Nonsynonymous#	Synonymous#	Nonsynonymous#	Synonymous [#]
<i>FOS</i> (2353)	28	6	3	6
<i>MVD</i> (4597)	9	29	2	2
<i>DEFB1</i> (1672)	0	2	7	15
<i>IFIT1</i> (3434)	13	5	15	11
<i>KLK6</i> (5653)	45	31	5	4

[#]Quantification of potencial new SNPs by SNP8. Quantification data was update in Jun, 26, 2006.

^fQuantification of annotated SNPs on dbSNP. Quantification data was update in Jun, 26, 2006 ().

^{*}Insertion/deletion polymorphisms were not considered.

FOS = v-fos FBJ murine osteosarcoma viral oncogene homolog Homo sapiens; MVD = mevalonate (diphospho) decarboxylase Homo sapiens; DEFB1 = defensin, beta 1 Homo sapiens; IFIT1 = interferon-induced protein with tetratricopeptide repeats 1 Homo sapiens; KLK6 = kallikrein 6 (neurosin, zyme) Homo sapiens.

CAPÍTULO V

CONSIDERAÇÕES FINAIS

A busca por polimorfismos em genes humanos tem sido descrita na literatura como fator importante no estudo da susceptibilidade a doenças virais (Ueda *et al.*, 2005; Lai *et al.*, 2005), a transmissão de doenças (Kostrikis *et al.*, 1999) e ao desenvolvimento de cânceres, como o de ovário e cólon retal (Fujita *et al.*, 2006). Através do programa SNP8 foi possível propor novos SNPs que ocorreram na seqüência codificadora da proteína (CDS) de genes humanos relacionados à patogenia do HPV.

Com o resultado obtido pelo SNP8 foi plausível perceber que a maioria dos SNPs já anotados no banco público internacional dbSNP (Sherry *et al.*, 1999) estão em regiões intrônicas, mostrando a importância da confirmação dos SNPs propostos pelo programa, os quais possibilitariam a complementação do dbSNP. Além do mais foi observado que o SNP8 propõe uma grande quantidade de novos SNPs não silenciosos, o que sugere uma investigação melhor desses polimorfismos visto que eles têm um potencial alto na alteração da estrutura da proteína e conseqüente modificação de sua atividade (Yamada *et al.*, 2005).

Dentre os 14 genes selecionados inicialmente para serem analisados pelo programa SNP8, três genes apresentaram melhor resultado quanto a quantidade de novos SNPs propostos e quanto a quantidade de ESTs que descreveram o mesmo SNP (Tabela 3). Para a maioria desses genes já foram descritos diferentes polimorfismos que vêm interferindo no processo infeccioso ocasionado pelo HPV

(Tabela 4), o que sugere posteriores análises quanto a sua eficácia como marcadores para a susceptibilidade ou não ao câncer de cólon uterino, como já vem sendo feito para outras doenças (Arraes, *et al.* 2004; Teles *et al.* 2005; de Souza *et al.* 2006; Segat *et al.* 2006). Finalizando, isto demonstra a aplicação do SNP8 na identificação de potenciais polimorfismos relacionados a características específicas tanto em humanos como em outras espécies de interesse para as ciências biomédicas, ciências biológicas, agricultura ou pesquisa.

Tabela 3 Caracterização do SNP melhor descrito para os genes que demonstraram melhor resultado após a análise pelo SNP8.

Gene (ID) ^a	Nucleotídeo ^b	SNP	EST(n) ^c	EST _{total} (n) ^d	%EST ^e	Tipo	Resíduo
<i>KLK7</i> (5650)	34	T/C	24	32	75,0	sinônima	Leucina
<i>SERPINB2</i> (5055)	279	G/A	17	56	30,36	sinônima	Alanina
<i>ALDH1A</i> (216)	225	C/T	94	216	43,52	sinônima	Serina

^aKLK7 - kallikrein 7 (chymotryptic, stratum corneum) [Homo sapiens]; SERPINB2 - serpin peptidase inhibitor, clade B (ovalbumin), member 2 [Homo sapiens]; ALDH1A1 - aldehyde dehydrogenase 1 family, member A1 [Homo sapiens].

^bDescreve a posição nucleotídica do SNP no CDS.

^cn é o número de ESTs que descrevem para o mesmo SNP.

^dn é o número total de ESTs que foram encontradas para esse descrever os SNPs para esse gene.

^eValor em porcentagem da quantidade de ESTs que descreveram o mesmo SNP.

Tabela 4 Relato de diferentes polimorfismos já descritos e relacionados à patogenia do HPV para os genes que tiveram melhor resultado após a análise pelo SNP8.

Gene (ID) ^a	Referência
<i>KLK7</i> (5650)	Santin <i>et al.</i> , 2004
<i>SERPINB2</i> (5055)	Riethdorf <i>et al.</i> , 1999; Darnell <i>et al.</i> , 2003; 2005
<i>ALDH1A</i> (216)	Santin <i>et al.</i> , 2004; Tian <i>et al.</i> , 2004; Talieri <i>et al.</i> , 2004

^aKLK7 - kallikrein 7 (chymotryptic, stratum corneum) [Homo sapiens]; SERPINB2 - serpin peptidase inhibitor, clade B (ovalbumin), member 2 [Homo sapiens]; ALDH1A1 - aldehyde dehydrogenase 1 family, member A1 [Homo sapiens].

REFERÊNCIAS

- Arraes,L.C. *et al.* (2004) MBL2 mother's genotype influences HIV-1 vertical transmissions in Brazilian children. *J. Brasileiro AIDS*, **5**, 193-232.
- Darnell,G.A. *et al.* (2003) Inhibition of retinoblastoma protein degradation by interaction with the serpin plasminogen activator inhibitor 2 via a novel consensus motif. *Mol. Cell Biol.*, **23**, 6520-6532.
- Darnell,G.A. *et al.* (2005) Silencing of integrated human papillomavirus type 18 oncogene transcription in cells expressing SerpinB2. *J. Virol.*, **79**, 4246-4256.
- de Souza,P.R.E. *et al.* (2006) CCR5 promoter polymorphisms and HIV-1 perinatal transmission in Brazilian children. *J. Reprod Immunol*, **69**, 77-84.
- Fujita,K. *et al.* (2006) Novel Single Nucleotide Polymorphism of UGT1A9 Gene in Japanese. *Drug Metab. Pharmacokinet.*, **21**, 79-81.
- Kostrikis,L.G. *et al.* (1999) A polymorphism in the regulatory region of the CC-chemokine receptor 5 gene influences perinatal transmission of human immunodeficiency virus type 1 to African-American infants. *J Virol*, **73**, 10264–10271.

Lai,H-C *et al.* (2005) Genetic polymorphisms of FAS and FASL (CD95/CD95L) gene in cervical carcinogenesis: an analysis of haplotype and gene-gene interaction. *Gynecol. Oncol.*, **99**, 113-118.

Riethdorf,L. *et al.* (1999) Urokinase gene expression indicates early invasive growth in squamous cell lesions of the uterine cervix. *J. Pathol.*, **189**, 245-250.

Santin,A.D. *et al.* (2004) The serine protease stratum corneum chymotryptic enzyme (kallikrein 7) is highly overexpressed in squamous cervical cancer cells. *Gynecol. Oncol.*, **94**, 283-288.

Segat L. *et al.* (2006). IL-18 gene promoter polymorphism is involved in HIV-1 infection in a Brazilian pediatric population. *Immunogenetics. In Press*.

Sherry,S.T. *et al.* (1999) dbSNP – Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res.*, **9**, 677-679.

Talieri,M. *et al.* (2004) Expression analysis of the human kallikrein 7 (KLK7) in breast tumors: a new potential biomarker for prognosis of breast carcinoma. *Thromb. Haemost.*, **91**, 180-186.

Teles,F.R. *et al.* (2005) Trends in dengue diagnosis. *Rev. Med. Virol.*, **15**, 287-302.

Tian,X. *et al.* (2004) Expression of human kallikrein 7 (hK7/SCCE) and its inhibitor antileukoprotease (ALP/SLPI) in uterine endocervical glands and in cervical adenocarcinomas. *Oncol. Rep.*, **12**, 1001-1006.

Ueda,M. *et al.* (2005) FA gene promoter –670 polymorphism (A/G) is associated with cervical carcinogenesis. *Gynecol. Oncol.*, **98**, 129-133.

Yamada,H. *et al.* (2005) Effect of splice-site polymorphisms of the TMPRSS4, NPHP4 and ORCTL4 genes on their mRNA expression. *J. Genet.*, **84**, 131-136.

ANEXOS

12.067 - Produtos Naturais

XIX Reunião Anual da Federação de Sociedades de Biologia Experimental

25 a 28 de agosto de 2004 | Aguas de Lindóia SP Brasil

CHITOSAN MICROPARTICLES FOR DELIVERY OF ANTIVIRAL AGENTS.¹

Camarotti-Lima, A. C.¹; ² Paulo, M. d. Q.; ³ Amorim, R. V.; ¹ Biologia Molecular, UFPE; ² Química, UFPB; ³ Biologia Molecular, UFPB; .

Objetivo:

Compare two methods (Met I and Met II) used for producing chitosan (CH)/polyethylene glycol (PEG) microparticles as the basis of controlled delivery system, *in vitro*, of the antiviral (ATV) agents.

Métodos e Resultados:

CH, with 90.3% degree of deacetylation, is a polysaccharide composed by 2-amino-2-deoxy- β -D-glucan. The ATV utilized was a coumarine derivative of the methanolic extract from *Coutarea hexandra* (Rubiaceae), a compound that generates inhibitory action against shingles. Microparticles of CH/PEG (4:1 w/w) were prepared by ionic cross-linking and precipitation with sodium sulfate. In Met I, the ATV was mixed with CH/PEG solution before the simultaneous cross-linking and precipitation. In Met II, the ATV was incubated with pre-formed microparticles for 48h. Through optical macro and microscopy no difference was observed from the use of any of methods. However, through infrared (IR) spectra the association between the compounds we used, seemed to be have differently according to the methods used. A noticeable difference was obtained from the encapsulation efficacy: 68.5% by using Met I and 98.6% by using Met II. However, delivering assays showed that in Met I the ATV delivered per minute was faster and 7 times greater than in Met II. The negatives controles to delivering assays was the microparticles without ATV. The action of pH 2.0 per 2 hours on both formulations showed that just a small amount of the drug was delivery in Met I and not any in Met II, although in both formulations the ATV maintained their integrities as observed in the controls.

Conclusões:

Our results suggest that the microaprticules of CH/PEG produced by Met I seems to be a very promising vehicle to controlled delivery system when administered orally. This formulation also showed that a large amount of the drug was delivered during the six hours of the assays in a buffer solution at pH 7.0 and that it was resistant to the acid pH of the stomach for at least two hours.

Apoio Financeiro: CNPq ,UFPE

PREMIADO

50.003 - Biologia Computacional

XX Reunião Anual da Federação de Sociedades de Biologia Experimental
24 a 27 de agosto de 2005 | Águas de Lindóia SP Brasil

IDENTIFICATION OF CANDIDATE SNPS FROM MICRO-ARRAYS DATABASE IN GENETIC PANELS AFFECTED BY HUMAN PAPILLOMAVIRUS .

¹ Camarotti-Lima, A. C. ^{**}; ² Santana, NA ^{*}; ³ Andrade, RS ^{*}; ⁴ Oliveira, JRM ; ⁵ Lima Filho, JL ; ^{1,2,3,4} Biologia Molecular, UFPE; ⁵ Biochemistry, UFPE; .

Objetivo:

We have been able to identify non previously reported Single Nucleotide Polymorphisms (SNPs), based on micro-arrays database of genes down/up regulated by HPV infection, such as the small-proline protein (SPRR) 2A and 2B, β -defensine 1 (DEFB1) and Interferon- α (IFI27).

Métodos e Resultados:

Initially we searched through microarray database of HPV studies in order to identify potential genes that were down regulated or upregulated after HPV infection. Then, we tried to identify candidate SNPs from ESTs database (NCBI, Goldenpath, etc) of these genes involved with the viral infection/progression. After that we built a panel of new candidate SNPs that are potential targets of physiopathology and drug action in HPV. Right now we are on the verge of validating our findings in micro-arrays of human genes.

The preliminary analyses identified potential new SNPs: 10 for SPRR2A, 14 for SPRR2B , 19 for DEFB1 and 33 for IFI27.

Conclusões:

Those results demonstrated the importance of these genes in studies of single nucleotide polymorphisms that could be associated with HPV infection/progression.

Apoio Financeiro: FINEP ,UFPE

ESTUDO DE POTENCIAIS NOVOS SNPs EM PACIENTES INFECTADAS PELO HPV.

**Santana, NA.¹, Camarotti-Lima, A. C.¹, Andrade, RS.¹, Oliveira, JRM.¹, Lima Filho, JL.²,
¹Biologia Molecular, UFPE; ²Bioquímica, UFPE.**

Além da determinação do tipo de papilomavírus humano (human papillomavirus – HPV) infectante na população, estudos atuais também vêm dando importância a polimorfismos de base única (Single Nucleotide Polymorphisms – SNPs) os quais constituem a forma mais comum de variação nas seqüências de DNA, sendo o seu estudo útil na identificação de muitas associações entre o genótipo e o fenótipo de cada indivíduo. O presente trabalho tem como objetivo a identificação de novos SNPs de genes que tiveram a sua regulação alterada em células infectadas pelo HPV. Através de estudos previamente realizados pela metodologia do micro-array, entre os anos de 2000 e 2004, foram selecionados os genes que se encontraram repetidos nesses trabalhos, são eles: β -defensina 1 (DEFB1), small-proline rich protein (SPRR) 1A, 2A e 2B, HLA-DRB5 e Interferon- α (IFI27). Primeiramente foram selecionadas seqüências de ESTs, Exons e SNPs já cadastradas nos banco de dados mundiais como o NCBI (Nacional Center for Biotechnology Information) e o Goldenpath, em seguida essas seqüências foram comparadas através do script computacional 7ClustalSNP. Com os resultados obtidos foi montado um painel molecular que sugeria potenciais novos SNPs relacionados à infecção pelo HPV para cada gene selecionado. As análises realizadas além de confirmarem os SNPs já cadastrados no NCBI (<http://www.ncbi.nlm.nih.gov>), identificaram potenciais novos SNPs, sendo 16 para DEFB1, 8 para SPRR2A, 6 para SPRR2B e mais de 100 para os genes IFI27 e HLA-DRB5. Os resultados demonstraram a importância desses genes nos estudos de polimorfismos que podem ser associados à infecção/progressão do HPV.

Apoio Financeiro: FINEP, CAPES, UFPE.

Resumo apresentado no XVII Encontro de Genética do Nordeste

Local: Mar Hotel, Recife – PE

Data: 12 – 16 de Abril de 2006.

PREMIADO