

Caracterização de *HIPK2*, um novo gene humano

Gaëlle Chopin Stukart

Dissertação apresentada à Fundação Antônio Prudente
para a obtenção do título de Mestre em Ciências.

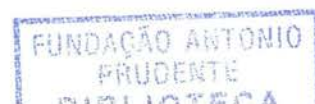
Área de concentração: Oncologia

Orientador: Dr. Emmanuel Dias Neto



São Paulo

2001



Stukart, Gaëlle Chopin

Caracterização de *HIPK2*, um novo gene humano / Gaëlle Chopin
Stukart. -- São Paulo, 2001.

95p.

Dissertação(Mestrado)–Fundação Antônio Prudente.

Curso de Pós-Graduação em Ciências-Área de concentração: Oncologia.

Orientador: Emmanuel Dias Neto.

Descritores: 1. HOMEODOMÍNIO. 2. GENES HOMEBOX. 3. FATORES DE TRANSCRIÇÃO. 4. PROTEÍNA QUINASE. 5. ISOFORMAS. 6. EXPRESSÃO GÊNICA. 7. ORGANOGÊNESE.

Dedico esta tese a meu eterno amor,
Henrique Parsons, que me apoiou e
incentivou durante todo este trabalho.
Obrigada por seu ombro sempre acolhedor,
sua paciência infinita, seu doce olhar e seu
carinho sincero.

Agradecimentos

Ao meu orientador, **Emmanuel Dias Neto**, pela liberdade que me conferiu, permitindo meu crescimento através da solução de problemas, e também pela inestimável ajuda naqueles nas quais somente sua experiência e gabarito puderam fornecer a saída segura.

A meu noivo **Henrique Parsons** por seu apoio, paciência e participação sempre.

A meus pais por todo o embasamento moral e educacional, e em especial, a minha mãe, **Françoise**, por sua compreensão e seu carinho e ao meu pai, **Gregorio**, pelo instinto perseverante.

A meus sogros, **Bernadete e Harry**, pelo grande interesse e incentivo.

Ao Prof. Dr. **Andrew Simpson** pela oportunidade de realizar este trabalho.

Ao **Instituto Ludwig**, através do Prof. Dr. **Ricardo Brentani**, pelo apoio e patrocínio, em conjunto com a **FAPESP**.

A meus professores da pós-graduação, Dr. **Luis Fernando Reis**, Dra. **Vilma**, Dr. **Roger Chammass**, Dra **Luisa Villa** e Dr. **Fernando Soares** pelo estímulo a pesquisa e ampliação de meus horizontes.

Ao Prof. Dr. **Fábio Nunes**, por sua colaboração dinâmica e pela produtiva troca de conhecimentos.

Ao Prof. Dr. **Luis Fernando Reis**, por seu apoio e seus conselhos.

Aos integrantes do grupo do Prof. Dr. Luis Fernando, **Adriana**, **Alex**, **Beatriz**, **Ricardo** e **Sibele**, pela fundamental, e jamais negada, ajuda.

À **Maria Cristina Costa**, por sua ajuda, principalmente no seqüenciamento das amostras dos diferentes animais.

À **Élida**, pelas frutíferas discussões científicas e todas as dúvidas tiradas.

À **Anamaria**, pelas discussões científicas e apoio.

À **Anna Cristina e Elisângela** pelo seqüenciamento.

À **Jane e a Elisângela** pelo apoio pessoal e técnico.

Aos alunos da Prof. Dra Mari Sogayar, **Diana, Christian, Regina Maki**,
pela ajuda pelas discussões.

Ao Prof. Dr. **Fabício Santos**, Prof. Dr. **Luiz Coutinho e José Garcia**
pelo material cedido.

À **Ana Maria** e à **Márcia** da secretaria da pós-graduação pela sua
dedicação e eficiência.

À bibliotecária **Suely**, pela revisão bibliográfica e ficha catalográfica.

Abreviaturas

aa – aminoácido

cDNA – DNA complementar

DNA – ácido desoxirubonucleico

dpc – dia(s) pós-coito

EST – *expressed sequence tag*

HID – domínio de interação com as homeoproteínas

HIPK – *homeodomain-interacting protein kinase*

Hipk2 – *homeodomain-interacting protein kinase 2* de camundongo

HIPK2 – *homeodomain-interacting protein kinase 2* humana

HMGI(Y) - *high mobility group I*

kb – kilobases

mRNA - RNA mensageiro

NLS – sinal de localização nuclear

ORESTES - open reading frame EST

pb – pares de base

PCR – reação em cadeia da polimerase

PEST – prolina (P), ácido glutâmico (E), serina (S) e treonina (T)

PKM – *Mx-interacting protein kinase*

RCF – força centrífuga relativa

RNA – ácido ribonucleíco

RT-PCR – reação em cadeia da polimerase após transcrição reversa

SAGE – *serial analysis of gene expression*

SRS – sinal de retenção no *nuclear speckle*

SUMO – *small ubiquitin-like modifier-1*

TA – temperatura ambiente

TNF – fator de necrose tumoral

Ubl – *Ubiquitin-like protein*

Resumo

Stukart GS. **Caracterização de *HIPK2*, um novo gene humano.** São Paulo; 2001.

[Tese de Mestrado – Fundação Antônio Prudente].

Diversas e complexas cascatas de modificações pós-traducionais são importantes na regulação da função celular. Neste projeto descrevemos uma nova isoforma do gene que codifica *HIPK2* (*homeodomain-interacting protein kinase 2*), um membro da família das proteínas quinases nucleares que agem como co-repressoras dos fatores de transcrição *homeobox*. Esta isoforma, quando comparada a sequência recentemente descrita de *HIPK2*, possui 27 aminoácidos a menos no domínio de interação com as homeoproteínas. A expressão das duas variantes deste novo gene humano foi analisada por RT-PCR. Um terceiro *splicing* foi encontrado e parcialmente confirmado por análises de EST e sequenciamento dos produtos de RT-PCR. Através de *Dot Blot*, *Northern Blot* e hibridização *in situ*, mostramos que *HIPK2* é expressa de forma diferencial em tecidos humanos e murinos. Uma maior expressão deste gene é observada no sistema nervoso central, coração e rim.

Summary

Stukart GS. **Caracterização de *HIPK2*, um novo gene humano.** [Characterization of *HIPK2*, a new human gene]. São Paulo (BR); 2001. [Tese de Mestrado – Fundação Antônio Prudente].

Complex cascades of post-translational modifications play important roles in the regulation of cellular function. We here describe a new isoform of the human homeodomain-interacting protein kinase-2 (*HIPK2*), a member of the family of nuclear protein kinases that act as co-repressors of homeobox transcription factors. When compared with the recently described human *HIPK2* sequence, we found that this variant differs due to the absence of 27 amino acids located in the homeoprotein-interacting domain. The expressions of these two new isoforms were analyzed by RT-PCR. A third splicing variant form was detected and partially confirmed by EST analysis and RT-PCR sequencing. By using dotblot, Northern Blot and *in situ* hybridization we show that *HIPK2* is differentially expressed in human and mouse tissues. A higher expression level is observed in the central nervous system, heart and kidney.

Índice

1. Introdução.....	1
2. Objetivos.....	11
3. Materiais e métodos.....	12
4. Resultados.....	31
5. Discussão.....	69
6. Conclusão.....	83
7. Perspectivas Futuras.....	87
8. Referências Bibliográficas.....	88
9. Anexos.....	95

Lista de Figuras e Gráficos

Figura 1 – Esquema da obtenção da sequência completa do cDNA de <i>HIPK2</i>	32
Figura 2 – Colônias obtidas pela metodologia de <i>Shot Gun</i>	32
Figura 3 – Sequência completa de nucleotídeos e aa de <i>HIPK2</i>	34
Figura 4 – Comparação do domínio serina-treonina quinase de <i>HIPK2</i>	36
Figura 5 – Comparação entre a sequência de aa de <i>Hipk2</i> e <i>HIPK2</i>	37
Figura 6 – Sequência genômica a partir da região de 1.259pb de <i>HIPK2</i>	41
Figura 7 – Desenho esquemático da distribuição de íntrons e éxons de <i>HIPK2</i>	41
Figura 8 – Comparação da sequência de aa das três isoformas de <i>HIPK2</i>	42
Figura 9 – Sequência genômica da região gera as duas isoformas de <i>HID</i>	43
Figura 10 – PCR da região codificadora das repetições de alaninas.....	46
Figura 11 – Sequências dos diferentes animais.....	47
Figura 12 – Sequência de aa de <i>HIPK2</i> comparada com a de outros animais.....	48
Figura 13 – Análise da expressão de <i>HIPK2</i> por <i>Northern Blot</i>	50
Figura 14 – Análise da expressão de <i>HIPK2</i> por <i>Dot Blot</i>	51
Figura 15 – Hibridização <i>in situ</i> de <i>HIPK2</i>	56
Figura 16 – PCR semi-quantitativa em estágios do desenvolvimento renal.....	59
Figura 17 – PCR em estágios do desenvolvimento de cérebro e coração.....	59
Figura 18 – Expressão das diferentes isoformas de <i>HIPK2</i> em tecidos humano.....	60
Figura 19 – Expressão das diferentes isoformas em LC de gliomas humanos.....	62
Figura 20 – Expressão das diferentes isoformas em LC de gliomas de ratos.....	62
Figura 21 – Desenho esquemático da região genômica amplificada por D e T.....	64

Figura 22 – Desenho das ESTs humanas similares aos introns I e III.....	65
Figura 23 – Análise do seqüenciamento amplificação com D e TI3.....	67
Gráficos 1 a 4 – Comparação das diferentes expressões obtidas pelo <i>Dot Blot</i>	52
Gráficos 5 e 6 – Comparação dos dados obtidos por SAGE.....	54

Lista de Tabelas

Tabela 1 – Tecidos utilizados neste projeto e sua origem.....	12
Tabela 2 – Iniciadores específicos de <i>HIPK2</i> utilizados no projeto.....	20
Tabela 3 – Iniciadores usados para a investigação do plimorfismo de YH.....	22
Tabela 4 – Iniciadores usados para a investigação da terceira isoforma.....	30
Tabela 5 – Organização parcial de éxons e íntrons deste gene.....	39
Tabela 6 – Animais utilizados no estudo da região polimorfica do domínio YH.....	45
Tabela 7 – Resultados da análise de SAGE.....	54

1. Introdução

1.1a. Fatores de transcrição

Os fatores de transcrição são proteínas que regulam a expressão gênica através de pelo menos dois domínios funcionais básicos, um de ligação ao DNA dos genes alvo e outro de transativação, que pode ativar ou reprimir a expressão destes genes, dependendo de seus domínios e da interação com outras proteínas reguladoras da transcrição. Vários fatores de transcrição já foram descritos em diferentes organismos, caracterizados pela presença de domínios específicos de ligação ao DNA com seqüências de aminoácidos (aa) e estruturas tridimensionais em comum (GEHRING et al. 1994a; URRUTIA 1997). Esta observação levou à classificação dos fatores de transcrição em diferentes famílias segundo o seu domínio de ligação ao DNA, entre elas a família dos genes *homeobox* (GEHRING et al. 1994a; URRUTIA 1997).

1.1b. Genes homeobox

Os genes da família *homeobox* têm fundamental importância nos processos de desenvolvimento, como molde de formação, organogênese e determinação do destino de cada célula em vários organismos (GEHRING et al. 1994a; KIM et al. 1998; KASAHARA e IZUMO 1999). Os genes *homeobox* foram descritos inicialmente em estudos de desenvolvimento realizados em drosófilas. Durante a embriogênese deste inseto, os produtos dos genes *homeobox* têm a função de estabelecer o plano axial do desenvolvimento do corpo e sua segmentação, determinando a identidade de cada segmento. Uma vez pronto o plano básico do corpo estas proteínas controlam a morfogênese de suas diferentes partes e a diferenciação celular (RUBENSTEIN e PUELLES 1994).

Os membros desta família de proteínas reguladoras (homeoproteínas) possuem uma sequência de 180 a 183 nt (sequência *homeobox*) que codifica 60 a 61 aa, o homeodomínio, que é responsável pela ligação da proteína ao DNA (CILLO et al. 1992; PETRINI et al. 1992; EDELMAN e JONES 1993; RUBENSTEIN e PUELLES 1994; TIBÉRIO et al. 1994; BIEBERICH et al. 1996; KASAHARA e IZUMO 1999). No curso da evolução, a sequência de aminoácidos do homeodomínio foi altamente conservada, sendo encontrada em metazoários, plantas e fungos (MARK et al. 1997; CILLO et al. 1999), indicando forte pressão evolutiva no sentido de preservar este domínio (McGINNIS e KRUMLAUF 1992; KRUMLAUF 1994; GEHRING et al. 1994a; MANN e CHAN 1996).

O homeodomínio possui uma estrutura tridimensional semelhante ao motivo hélice-volta-hélice encontrado em proteínas regulatórias de procariotos (McGINNIS e KRUMLAUF 1992; EDELMAN e JONES 1993; RUBENSTEIN e PUELLES 1994; GEHRING et al. 1994b; TENEN et al. 1997). Este motivo constitui de três α -hélices, onde as hélices 2 e 3 formam o motivo hélice-volta-hélice (VOLLMER e CLERC 1998). A terceira hélice, na porção carboxi-terminal da proteína, é a responsável pela ligação à grande curvatura do DNA enquanto a primeira hélice liga-se na pequena curvatura (GEHRING et al. 1994b; VOLLMER e CLERC 1998).

Contrariamente às funções biológicas altamente específicas *in vivo* das homeoproteínas, estudos de ligação dos homeodínios *in vitro* demonstraram que a maioria se liga a uma sequência específica, a sequência 5' TAAT 3', denominada em inglês *core motif* (GEHRING et al. 1994b; PATTERSON et al. 1998; KASAHARA e IZUMO 1999). Este aparente contraste sugere que a especificidade do gene alvo de cada homeoproteína *in vivo* pode ser conseguida pela combinação dos diferentes fatos:

- (a) A combinação de interações com outros cofatores (CHAN et al. 1994; VAN DIJK e MURRE 1994; MANN e CHAN 1996), que são, na maioria dos casos, fatores de transcrição com domínios próprios de ligação ao DNA (CHAN et al. 1994, GEHRING et al. 1994a; KIM et al. 1998), que direcionam as diferentes homeoproteínas para seus diferentes genes-alvo, regulando-os temporal e espacialmente (KIM et al. 1998; KRUMLAUF 1994; MCGINNIS e KRUMLAUF 1992);
- (b) pequenas diferenças de afinidade entre as homeoproteínas e seus sítios-alvo;
- (c) modificações pós-traducionais como a fosforilação ou não da proteína (GEHRING et al. 1994b; VOLLMER e CLERC 1998; KASAHARA e IZUMO 1999).

Muitos fatores de transcrição parecem ser regulados por fosforilação e desfosforilação. Várias proteínas com homeodomínios mostraram-se fosforiladas em diferentes sítios por diferentes proteínas-quinase, como a proteína humana *Cut*, que apresenta homeodomínio e é regulada por fosforilação através de proteínas-quinase C e de quinases caseínas 2 (KASAHARA e IZUMO 1999).

Os genes *homeobox* foram divididos em duas categorias gerais: o grupo HOX (ortólogo do grupo HOM-C da drosófila) e o grupo órfão, ou não agrupado. O primeiro, envolvido principalmente na divisão axial do corpo, encontra-se agrupado em 4 *clusters*. O segundo encontra-se disperso no genoma e não parece estar envolvido com a divisão axial do corpo e sim com a organogênese (PATTERSON et al. 1998).

1.1c. Genes homeobox e doenças congênitas

Devido ao papel crucial dos genes *homeobox* no desenvolvimento, já foi descrito o envolvimento de alterações destes em algumas doenças congênitas. No entanto ainda há muito a se investigar quanto ao envolvimento desta complexa família de genes em doenças congênitas.

Alguns exemplos de alterações em produtos de genes *homeobox* que levam a doenças congênitas são citados a seguir. A proteína HOXD13 está relacionada à síndrome humana de sinpolidactilia (INNIS 1997; MARK et al. 1997). Esta proteína possui uma pequena repetição de alaninas na região N-terminal que permite sua ligação ao DNA, mas inativa a sua função de transativação de outras proteínas (MARK et al. 1997). Na síndrome de Waardenburg, responsável por aproximadamente 2% de todos os casos de surdez congênita, o gene *homeobox PAX3* possui mutações que levam à perda da função de sua proteína (MARK et al. 1997). Mutações no gene *homeobox NKX2-5* já foram extensamente caracterizadas em doenças congênitas cardíacas (SCHOTT et al. 1998; BENSON et al. 1999; PABST et al. 1999, TANAKA et al. 1999; BIBEN et al. 2000).

1.1d. Genes homeobox e câncer

Os genes *homeobox* compreendem uma vasta coleção de genes reguladores do desenvolvimento de embriões, envolvendo desde invertebrados até mamíferos. Verificou-se também a presença de seus transcritos em tecidos adultos, tais como rim e células da linhagem hematopoética (PETRINI et al. 1992). É de grande importância estudar e avaliar cuidadosamente a relação entre os genes *homeobox* e os mecanismos envolvidos na oncogênese, visto que ambos estão envolvidos com a aquisição, regulação e manutenção da identidade celular (CILLO et al. 1999). Expressões anormais de vários genes *homeobox* foram descritas em várias formas de câncer, incluindo o pulmonar (TIBERIO et al. 1994),

renal (CILLO et al. 1992), de cólon (DE VITA et al. 1993) e de mama (FRIEDMANN et al. 1994). Adicionalmente, mutações, deleções e translocações que afetam genes *homeobox* foram extensamente caracterizadas em outras formas malignas tais como as leucemias (MOSKOW et al. 1995).

Alguns exemplos de genes *homeobox* envolvidos com neoplasias são descritos a seguir. O gene *Hoxa-1*, em ratas, não é expresso nas glândulas mamárias normais porém nas neoplasias há presença do transcrito (FRIEDMANN et al. 1994). A superexpressão deste gene em cultura de células da linhagem NIH-3T3 ou 208-F leva à transformação celular, permitindo que estas cresçam e formem focos nas placas de ágar. Injetando estas células em ratos atímicos estes desenvolvem um tumor no local da inoculação. Este e outros estudos correlatos confirmam que alguns genes *homeobox* (como *Hoxa-1*), devem constituir uma nova família de oncogenes (FRIEDMANN et al. 1994).

Outro exemplo da expressão alterada de genes *homeobox* foi encontrado em humanos. Os genes *HOX-2A* e *HOX-2E* são expressos ativamente em rins normais, mas estão ausentes em biópsias de tumores enquanto o gene *HOX-3H*, que não é expresso no rim normal, é expresso em carcinomas renais (ANBAZHAGAN e RAMAN 1997). Estudos de quantificação da expressão gênica mostraram que alguns genes *homeobox* podem constituir uma família de oncogenes enquanto que alguns outros devem constituir uma família de genes supressores de tumor.

Diferenças no tamanho do transcrito também foram observadas. Em ratos, infecções por retrovirus podem provocar leucemias e em conjunto há a alteração da expressão do gene *homeobox Meis1* pela integração do vírus neste. O resultado obtido neste trabalho com o gene *Meis1* mostra que diferentes tamanhos do transcrito são visualizados por *Northern Blot*, provavelmente devido a um *splicing* alternativo (MOSKOW et al. 1995). As formas truncadas do transcrito podem levar à perda da região 3' não codificadora,

alterando a estabilidade do mRNA e contribuindo para a transformação das células hematopoéticas (Moskow et al, 1995). A forma truncada também pode criar uma proteína com funções distintas da original. Os autores concluíram que a integração viral no *Meis1* resulta na desregulação da proteína que pode levar ao bloqueio da diferenciação ou a um aumento na proliferação das células mielóides em desenvolvimento. Esta alteração na via de diferenciação colabora com a formação da leucemia mielóide (MOSKOW et al. 1995). As modificações no tamanho do transcrito também foram vistas em neoplasias renais (CILLO et al. 1992).

O gene *homeobox NKX3.1*, que se encontra no cromossomo 8p21, tem importante papel no desenvolvimento normal da próstata e na manutenção do fenótipo no órgão adulto. Em câncer de próstata esta região cromossômica é geralmente deletada. Estes dados apontam um possível envolvimento deste gene na desdiferenciação celular durante o câncer de próstata (HE et al. 1997).

1.1e. A família de genes homeobox NK

A família dos genes homeóticos *NK* foi descrita em drosófila por KIM e NIREMBERG (1989), através da varredura de bibliotecas de DNA com oligonucleotídeos degenerados das seqüências codificadoras dos homeodomínios. Quatro genes foram identificados (*NK1* a *NK4*) e posteriormente classificados em duas classes, segundo a similaridade observada entre as seqüências de aminoácidos, a classe NK-1, contendo o gene *NK1*, e NK-2, contendo os genes *NK2*, *NK3* e *NK4* (HARVEY 1996). Os genes da classe NK-2 também foram denominados em drosófila pelo fenótipo que suas proteínas mutadas causavam, sendo *NK2* também denominado *vnd* (*ventral nervous system defective*), *NK3*, *bap* (*bagpipe*) e *NK4* denominado também de *tin* (*tinman*) (HARVEY 1996).

As proteínas da família NK, diferentemente das outras homeoproteínas que se ligam à sequência de DNA com o motivo 5'TAAT 3', se ligam a seqüências de DNA com o motivo 5'(C/T)AAG 3'. A classe NK-2 possui além da seqüência de aminoácidos do homeodomínio, uma pequena seqüência de 10 aminoácidos na região amino-terminal da proteína, denominada domínio TN (HARVEY 1996; PATTERSON et al. 1998). Este domínio é similar à parte central (*core*) do domínio do elemento repressor eh-1 (PATTERSON et al. 1998).

Os genes *NK* estão envolvidos na especificação do tipo celular e manutenção de fenótipos diferenciados em vários tecidos de organismos evolucionariamente distantes (PABST et al. 1999). A participação desta família de genes homeóticos foi extensamente caracterizada no coração dos vertebrados, principalmente os genes relacionados a *tinman* de drosófila, sendo importantes para a especificação das células precursoras cardíacas e para a morfogênese deste órgão (DOEVENDANS e BILSEN 1996; HARVEY 1996; PATTERSON et al. 1998; SCHWARTZ e OLSON 1999; TANAKA et al. 1999). Entre estes genes podemos citar os identificados em vertebrados, *Nkx2-5*, *Nkx2-3*, *Nkx2-6*, *Nkx2-7*, *Nkx2-8* e *Nkx2-9* (PATTERSON et al. 1998). Camundongos nocautes para o gene *Nkx2-5* morriam intra-útero com 9 dias pós-coito (dpc) com o tubo cardíaco ainda linear, ou seja, sem a formação das câmaras, mostrando a importância deste gene na organização regional deste órgão (DOEVENDANS e BILSEN 1996). Mutações no gene humano *NKX2-5* foram relacionadas a doenças cardíacas congênitas, como defeitos nos septos ventriculares ou atriais e tetralogia de Fallot (BENSON et al. 1999).

Genes da família NK também foram estudados no desenvolvimento de várias regiões do cérebro, entre eles *Nkx2-1*, *Nkx2-2*, *Nkx5-1*, *Nkx5-2* e *Nkx 1-2* ou *Sax-1*, que são importantes para diferenciação regional específica das áreas do cérebro (VOLLMER e CLERC 1998). Esta família também está expressa durante o desenvolvimento de outros

órgãos como faringe, tireóide, estômago, baço, intestino delgado, próstata e pâncreas (HE et al. 1997; BHATIA-GAUR et al. 1999; PABST et al. 1999; SCHWARTZ e OLSON 1999; SANDER et al. 2000). O gene *NKX3-1* e seu ortólogo em camundongo estão envolvidos na carcinogênese de próstata, sendo possíveis candidatos a genes supressores de tumor, que também estão envolvidos no desenvolvimento deste órgão (BHATIA-GAUR et al. 1999; BOWEN et al. 2000).

1.1f. Proteína quinase 2 com domínio de ligação a homeoproteínas

Cofatores de homeoproteínas auxiliam estas proteínas na sua especificidade, ou seja, facilitam o direcionamento e função destas nos seus genes-alvo (BHATIA-GAUR et al. 1999; BOWEN et al. 2000). Em 1998, Kim e colaboradores descreveram em camundongos uma nova família de cofatores de homeoproteínas, a família HIPK (*Homeodomain-Interacting Protein Kinase*), que diferentemente dos outros cofatores já descritos para as proteínas codificadas por genes *homeobox*, possuíam atividade quinásica. HIPKs potencializam as atividades transcricionais, mais especificamente a atividade repressora, das homeoproteínas da família NK. Na descoberta desta nova família, os autores estavam à procura de proteínas que interagissem com a família NK utilizando, no ensaio de duplo híbrido, o *Nkx1.2* de camundongo, membro da família NK1. Assim foram descritos os genes que codificam as proteínas Hipk1 e Hipk2 e posteriormente, em uma varredura de biblioteca de cDNA murino, o gene *Hipk3*. Estas três proteínas possuíam em comum o domínio quinase, o domínio de interação com as homeoproteínas, HID (do inglês *Homeoproteins-Interaction Domain*), a sequência PEST (prolina (P), ácido glutâmico (E), serina (S) e treonina (T)) descrita por Rogers e colaboradores, 1986, e uma região enriquecida com resíduos de histidina e tirosina, o domínio YH, localizado na região C-terminal de HIPKs (KIM et al. 1998).

Os domínios HID das HIPKs possuem pequenas diferenças entre si, tendo o HID de Hipk1 uma sequência adicional de aminoácidos (aa) e 75% de identidade com o HID de Hipk2 e Hipk3 72% de identidade com Hipk2 (KIM et al. 1998). As HIPKs interagem de maneira diferente com as homeoproteínas da família NK e Hox. Em particular Hipk2, entre as homeoproteínas testadas por Kim e colaboradores (1998), interage melhor com Nkx1.2, Nkx2.5 (ambas proteínas de camundongo), NK-1 e NK-3 e com menor intensidade com HoxC4 e HoxD4. Tratando-se de sua atividade quinásica, sabe-se que Hipk2 fosforila Nkx1.2, NK-1 e NK-3, porém desconhece-se sua atividade em Nkx2.5, HoxC4 e HoxD4 (KIM et al. 1998). Estudos de deleção em Hipk2 identificaram um domínio de co-repressão, mas para este ser completamente funcional é necessária a atividade quinásica.

A localização de Hipk2 é nuclear, especificamente nos *nuclear speckles* ou *dots*, compartimentos nucleares diferentes dos que contêm os fatores de *splicing* (KIM et al. 1998, 1999). Sinais de localização nuclear, NLSs (*Nuclear Localization Signals*) e um sinal de retenção nos *nuclear speckles*, SRS (*Nuclear Speckle Retention Signal*) foram identificados por técnicas de deleção (Kim et al., 1998 e Kim et al., 1999). Kim e colaboradores mostraram em 1999 que sua compartimentalização dentro do núcleo está relacionada às modificações pós-traducionais feitas por SUMO-1 (*Small Ubiquitin-like Modifier-1*), uma proteína parecida com ubiquitina. As UbIs (*Ubiquitin-like proteins*), diferente das ubiquitinas, não estão envolvidas na degradação de proteínas, mas sim na determinação da localização subcelular ou no aumento de estabilidade (KIM et al 1999).

1.2. Descoberta de novos genes humanos e seus possíveis *splicings*

Após a publicação da sequência nucleotídica do genoma humano nas revistas científicas Nature e Science em fevereiro de 2001, o novo desafio da pesquisa genômica passou a ser a identificação de todos os genes expressos, conhecendo seus diferentes

transcritos e funções no contexto celular. No início deste projeto, utilizando dados do Projeto Genoma Humano do Câncer, encontramos uma EST (do inglês, *Expression Sequence Tag*, ou etiqueta de sequência expressa) similar ao gene que codifica Hipk2, que havia sido descrito apenas em camundongos e em hamsters, neste denominado PKM (*Mx-interacting protein kinase*), descrito por Trost e colaboradores em 2000.

O Projeto Genoma Humano do Câncer, fruto de uma cooperação entre a FAPESP e o Instituto Ludwig de Pesquisas sobre o Câncer, gerou mais de um milhão de seqüências parciais de genes humanos, a partir de amostras biológicas derivadas principalmente de neoplasias. O sucesso deste projeto dependeu do uso de uma nova metodologia, criada e patenteada pelo Instituto Ludwig, chamada ORESTES (do inglês, *Open Reading Frames EST Sequence*). ORESTES permite seqüenciar a área central dos genes expressos (DIAS NETO et al. 2000), o que é totalmente complementar aos dados obtidos por técnicas tradicionais de seqüenciamento de bibliotecas de cDNA, onde se explora basicamente as extremidades dos genes. A importância da porção central é que esta representa a parte codificadora do gene, sendo possível assim obter dados relativos à seqüência das proteínas geradas, favorecendo o encontro de similaridades com proteínas já conhecidas, sejam elas humanas ou não. Além disto, ORESTES favorece o encontro de ESTs derivadas de genes pouco expressos, permitindo o estudo de uma gama de genes desconhecidos ou pouco conhecido em relação as ESTs.

Desta forma, neste projeto nos concentramos na caracterização de *HIPK2*, um novo gene humano, inicialmente identificado por similaridade entre uma ORESTES e o gene murino *Hipk2*.

2. Objetivos

Os objetivos deste trabalho foram caracterizar um novo gene humano denominado

HIPK2. Esta caracterização consistiu em:

- Obtenção do cDNA completo de *HIPK2* e análise de sua sequência genômica;
- Análise de seus diferentes domínios;
- Análise de suas diferentes isoformas;
- Análise de sua expressão em tecido humanos, de camundongos e de ratos por *Dot Blot*, *Northern Blot*, hibridização *in situ* e RT-PCR.

3. Materiais e métodos

3.1. Amostras biológicas

A tabela 1 indica os tecidos usados neste projeto e sua origem.

Tecido ou Linhagem Celular	Cedido por:
Cérebro, língua, cólon, mama, amostras normais e tumorais	Banco do Instituto Ludwig
Células do líquido amniótico	Banco do Instituto Ludwig
Ovário tumoral	Banco do Instituto Ludwig
Testículo tumoral	Banco do Instituto Ludwig
Tireóide tumoral	Beatriz Stolf (Inst. Ludwig)
Linhagens de células de carcinoma de cólon	Banco do Instituto Ludwig
Tecido de camundongos fetais, jovens e adultos	Prof. Dr. Fábio Daumas Nunes, laboratório de Biologia Molecular do Instituto de Odontologia da USP
Linhagens celulares de gliomas humanos A172 e T98G	Laboratório da Profa. Dra. Mari Sogayar, Depto de Bioquímica do Instituto de Química da USP
Linhagens celulares de gliomas de ratos ST1 e P7	Laboratório da Profa. Dra. Mari Sogayar, Depto de Bioquímica do Instituto de Química da USP
Tamanduá-bandeira, tamanduá-mirim, gato-do-mato, cachorro-do-mato	José Eduardo Garcia e Profa Eucleia Primo Betioli Contel, Depto de Genética da Faculdade de Medicina de Ribeirão Preto, SP
Preguiça comum, gambá, mico-leão, cavalo, cachorro,	Prof. Dr. Fabricio Rodrigues dos Santos, Depto de Biologia Geral do Inst. De Ciências Biológicas da UFMG
Galinha	Prof. Dr. Luiz L. Coutinho, ESALQ/USP, Piracicaba, SP
Porco, boi	Açougue

Tabela 1: *Tecidos utilizados neste projeto e sua origem.*

3.2. Obtenção da sequência completa de HIPK2

3.2a. Clones adquiridos comercialmente

Após uma busca no banco de ESTs do *GenBank*, encontrou-se 4 clones similares ao cDNA de *Hipk2*. Estes clones (números de acesso no *GenBank*: AW044305, AI580907, H85811 e AA4553887), foram adquiridos junto ao consórcio IMAGE através do grupo Alemão RZDP (*Deutsches Ressourcenzentrum für Genomforschung GmbH*). As bactérias contendo estes clones foram plaqueadas em LB sólido com ampicilina, e, após seleção por PCR das colônias que continham o inserto (metodologia descrita no tópico 3.2b), foi realizada uma *miniprep* (metodologia descrita no tópico 3.2b) e os plasmídeos foram seqüenciados com os iniciadores *M13 forward* e *M13 reverse* e para o clone H85811 também com o iniciador T3. Estes clones foram crescidos em 15ml de LB líquido (SAMBROOK et al. 1989) com 7,5µl de ampicilina (200mg/ml) e incubados sob agitação a 37°C, durante cerca de 18h. A *miniprep* foi realizada com o *kit Wizard Miniprep* (Promega) e o produto visualizado em gel de agarose 1% corado com brometo de etídio para posterior seqüenciamento. Dez microgramas dos plasmídeos com os clones H85811 e AA4553887 foram digeridos por 4 horas a 37°C com 20U das enzimas EcoRI e BamHI (New England Biolabs) para o clone H85811 e EcoRI e HindIII (New England Biolabs) para o clone AA4553887. O produto da digestão foi visualizado em gel de agarose 0,8% corado com brometo de etídio e as bandas foram eluídas do gel pelo *kit QIAquick™ Gel Extraction* (Quiagen) para serem posteriormente usadas como molde para a reação de *Shot Gun* PCR e, no caso do clone AA4553887, ser usado como sonda para os *Northern* e *Dot Blot*.

3.2b. *Shot Gun* PCR

Para obtermos seqüências derivadas de porções mais internas deste fragmento, trabalhamos em conjunto com o estudante Gabriel Mattos, utilizando um protocolo alternativo, desenvolvido em nosso laboratório, baseado em *Shot Gun* por PCR. A reação de *Shot Gun* PCR utilizou os iniciadores PBT137A (5' AATATATCCGGCCNNNCC 3', onde "N" representa qualquer um dos quatro nucleotídeos) e PBT052A (5' AAGATGGCAAAGCCTCCNNCCCNG 3') para a síntese das bibliotecas dos clones do Image H85811 e AA4553887 e os iniciadores DEG1 (5'AANGTNGGNCCCATBC 3', com "B" representando C, G ou T) e DEG2 (5'CCSCCSAGCCAVSRSCCSCCSCC 3', com "S" representando G ou C, "V" representando A, C ou G e "R" representando A ou G). Para a reação de amplificação utilizamos um volume final de 25µl, sendo 2,5µl 10X *PCR buffer* (GibcoBRL), 2,5µl MgCl₂ 50mM (GibcoBRL), 4,0µl dNTP 1,25mM (GibcoBRL), 1µl iniciador 25µM, 0,2µl *Taq polimerase* 5U/µl (GibcoBRL) e aproximadamente 2pg de DNA como molde. A ciclagem utilizada foi: 95°C por 2 minutos, 5 ciclos de 95°C por 45 segundos, 30°C por 1,5 minuto e 72°C por 1 minuto, 35 ciclos de 95°C por 45 segundos, 45°C por 1,5 minuto e 72°C por 1 minuto e finalizada a 72°C por 5 minutos. Para cada molde de DNA foram realizadas 3 reações de amplificação, uma com cada um dos iniciadores e uma terceira com os dois iniciadores simultaneamente. A ciclagem foi realizada no termociclador *PTC-100TM Programable Thermal Controller*, MJ Research, Inc., ou no termociclador *GeneAmp[®] PCR System 9700*, PE Applied Biosystems. Os produtos das três reações foram visualizados em gel de acrilamida 6%. Aliquotas das três reações foram então misturadas e aplicadas em um gel de agarose 0,8%, onde foi feita uma seleção das bandas de tamanho maior que 300pb. Os fragmentos amplificados foram eluídos da agarose pelo *kit QIAquickTM Gel Extraction* (Quiagen) para serem clonados e seqüenciados com os iniciadores específicos do plasmídeo.

3.2c. Visualização do produto da PCR em gel de poli-acrilamida

A reação de amplificação foi visualizada em um gel de poli-acrilamida 6% em tampão TBE 1X pH 8. (0,09M Tris-Borato e 0,002M EDTA). 3µl da reação foram aplicados ao gel com 1µl de tampão 6X (0,25% de Bromofenol 5%, 0,25% de Xileno-cianol 5% e 15% de Ficoll). O marcador de peso molecular utilizado foi o 100pb *ladder* (New England BioLabs ou GibcoBRL) a 40ng/µl, usando-se 400ng por gel. Após a separação das bandas por eletroforese, o gel foi corado para visualização. Para coloração o gel foi primeiramente incubado sob agitação continua por 5 minutos em 150ml da solução fixadora (0,5% de ácido acético glacial e 10% de etanol absoluto) a temperatura ambiente (TA) e em seguida foi acrescentado 2ml da solução de nitrato de prata 0,2% e a solução incubada por mais 5 minutos. Após o tempo de impregnação do DNA pela prata, lavou-se o gel rapidamente com água MilliQ e adicionou-se 150ml da solução reveladora (30% de NaOH e 0,5% de formaldeído) deixando também em agitação continua por 5 a 10 minutos, até a coloração atingir a intensidade ideal. Trocou-se a solução reveladora por uma nova, fixadora (150ml), e o gel foi seco em uma placa de vidro envolta em papel celofane úmido e outra camada de celofane úmido sobre o gel, evitando-se bolhas de ar entre o gel e o celofane, e deixado à TA durante cerca de 18h. Este protocolo de coloração é baseado no descrito por Sanguinetti e colaboradores, 1994.

3.2d. Clonagem

I. Ligação de insertos ao plasmídeo e transformação das bactérias

Os fragmentos foram clonados em pUC 18 SmaI/BAP com o *kit SureClonetm Ligation* (Amersham Pharmacia Biotech), segundo instruções do fabricante. A 100µl de bactérias JM109 quimiocompetentes foram adicionados 5µl da ligação do DNA com o plasmídeo e incubados em gelo por 30 minutos. Estes tubos foram colocados a 42°C por 1,5 minuto e novamente em gelo por 3 minutos. Foram então adicionados 500 µl de LB e as bactérias foram deixadas em recuperação do choque térmico por 45 minutos a 37°C. O volume total foi então plaqueado em LB sólido com ampicilina, X-Gal e IPTG (SAMBROOK et al. 1989).

II. PCR de colônia

Cada colônia foi numerada e coletada individualmente com um palito em um tubo *eppendorf* 0,2ml para a amplificação juntamente com 5pmol dos iniciadores *M13 forward* e *M13 reverse*, 2mM de MgCl₂, 0,25mM de dNTP, 1µl de 10X *PCR Buffer* e 0,5 unidades de Taq polimerase (GibcoBRL). A amplificação foi realizada no termociclador *GeneAmp®PCR System 9700* (PE Applied Biosystems,USA) e a ciclagem utilizada foi 94°C por 5 minutos, 35 ciclos de 94°C por 45 segundos, 55°C por 45 segundos e 72°C por 45 segundos e, ao final de 35 ciclos, 7 minutos a 72°C. Os produtos de PCR foram visualizados em gel de acrilamida 6% ou agarose 1%.

III. Purificação do plasmídeo (Mini preparação por lise alcalina - Miniprep)

As colônias que apresentaram inserto após PCR de colônia foram repicadas em 5ml de LB (SAMBROOK et al. 1989) com 450µg de ampicilina e incubados sob agitação a 37°C,

durante cerca de 18h. Os plasmídeos foram purificados com o *kit* Wizard Miniprep (Promega) e os produtos visualizados em gel de agarose 1% corado com brometo de etídio.

3.2e. Seqüenciamento

A reação de seqüenciamento foi realizada com o *kit* DYEnamicTM ET Terminator Cycle Sequencing (Amersham Pharmacia Biotech), usando 2µl do *mix* de seqüenciamento, 3,2 a 8pmols de um dos iniciadores e 3-10ng do produto de PCR ou 300-500ng de plasmídeo. Foi feita uma reação a 95°C por 20s e 25 ciclos de 55 ou 60°C (de acordo com a temperatura de anelamento de cada iniciador) por 15s e 60°C por 1 minuto em um termociclador GeneAmp® PCR System 9700, PE Applied Biosystems. Os produtos da reação foram então precipitados com 1µl de acetato de sódio fornecido com o *kit* e 40µl de etanol 95%, agitados rapidamente com o auxílio do vortex e incubados por 15 minutos à TA em local escuro. Após centrifugação a 15.300 RCF por 15 minutos, o sobrenadante foi aspirado a vácuo. Foram adicionados 250 µl de etanol 70% e centrifugou-se por mais 5 minutos a 15.300 RCF e novamente o sobrenadante foi aspirado e o precipitado secou por 5 a 10 minutos na estufa a 54°C. A este foi então adicionado 3µl de solução de ressuspensão fornecida pelo *kit* e foi aplicado 2µl no gel de seqüenciamento no seqüenciador ABI PRISMTM 377 DNA Sequencer.

3.2f. Extração de RNA total

O RNA total foi utilizado no preparo de membranas de *Northern Blot* ou nas RT-PCRs. A sua extração foi feita com TRIzol® Reagent (GibcoBRL), segundo as instruções do fabricante. Cinquenta a cem miligramas de tecido foram homogeneizados com 1ml de TRIzol e colocado em repouso por 5 minutos à TA. Foi então adicionado 0,2ml de clorofórmio ao tubo e este agitado com auxílio do vortex por 15 segundos. O tubo foi

novamente colocado em repouso por 3 minutos à TA e centrifugado em 12.000 RCF por 15 minutos a 4°C. A fase aquosa foi transferida para um novo tubo para a precipitação com 0,5ml de propanol e este incubado à TA por 10 minutos e então centrifugado em 12.000 RCF por 10 minutos a 4°C. O sobrenadante foi descartado e o precipitado de RNA total foi lavado com 1ml de etanol 75% gelado, misturado e centrifugado em 7.500 RCF a 4°C. O sobrenadante foi descartado e o precipitado foi seco à TA por 5 a 10 minutos e ressuspensionado em até 50µl de água DEPC. O RNA foi quantificado no espectrofotômetro *Ultrospec®2000* (Amersham Pharmacia Biotech) e estocado a -70°C. O RNA utilizado para a RT-PCR foi tratado com DNase (Promega) por 60 min a 37°C.

3.2f. Reverse Transcriptase-PCR (RT-PCR)

Para a obtenção da sequência completa do cDNA de *HIPK2* foram feitas RT-PCR com iniciadores específicos desenhados a partir dos dados obtidos com os dois clones do IMAGE, *Shot Gun* PCR e ESTs humanas similares a *Hipk2* encontradas no *GenBank*.

A síntese de cDNA foi feita com 1µl de RNA total derivado de cérebro humano (lobo frontal) e também com RNA total derivado de glioblastoma humano. A quantidade de RNA total utilizado para a síntese variou de 1 a 3µg. A síntese foi feita utilizando *Ready To Go RT Beads* (Amersham Pharmacia Biotech) e 0,5µg do iniciador pd(T)₁₂₋₁₈ ou 0,2µg pd(N)₆ durante 2 horas a 37°C. O produto foi estocado a -70°C. Para a síntese utilizou-se a enzima *Superscript II Reverse Transcriptase* (GibcoBRL) e 0,5µg do iniciador pd(T)₁₂₋₁₈ ou 0,2µg pd(N)₆ durante 1 hora a 42°C, segundo instruções do fabricante. O cDNA resultante foi estocado a -70°C.

Para a reação de amplificação, o cDNA estoque derivado de RNA de cérebro normal foi diluído em uma proporção de 1:10 e 1µl foi usado na PCR junto com 2,5pmol de cada

iniciador específico. Diversos iniciadores foram desenhados para a obtenção e análise de várias regiões deste gene. A tabela 2 lista a sequência completa e a posição no cDNA de HIPK2. A PCR foi feita com a enzima *Platinum®Taq DNA Polymerase High Fidelity* (GibcoBRL). Para esta reação utilizamos 2mM de MgSO₄, 0,25mM de dNTP, 1µl de 10X *High Fidelity PCR Buffer*, 0,5 unidades da enzima, 2,5pmol de cada iniciador específico e 1µl do cDNA diluído 1:10. A amplificação foi realizada no termociclador *GeneAmp®PCR System 9700* (PE Applied Biosystems, USA) e a ciclagem utilizada foi 94°C por 5 minutos, 40 ciclos de 94°C por 45 segundos, 56-62°C (segundo a temperatura de anelamento calculada para os iniciadores utilizados) por 45 segundos e/ou 68°C por 1-5 minutos, segundo o tamanho do fragmento a ser amplificado e ao fim destes 40 ciclos, 7 minutos a 72°C. Após o término da reação os tubos foram mantidos a 4°C. Alternativamente algumas reações foram feitas utilizando o *PCR Supermix* (GibcoBRL).

Iniciador	Sequência 5'-3'	Posição no cDNA de HIPK2
Homeo A	TATGTAAGGGTACTGGTTGAC	3668-3648
Homeo B	TCAGCGTCATCACCATCAGC	2728-2747
Homeo C	ACCGGTCAATAGTGCAGGC	2393-2375
Homeo D	TGCAAGCGTCGGGTGAATA	1806-1824
Homeo E	TCTCTGCTTCATGGTCATCT	1527-1508
Homeo F	AGACACAGGCTCAAGATGG	52-70
Homeo G	ATGTTCTCTTCTCTGCTCTAC	3910-3890
Homeo H	AGCACTGTGAGCCTCCTTG	507-525
Homeo I	CAAAAATGTGGACTCAAG	534-553
Homeo J	AACCGTCAGCACCTCCTTG	350-368
Homeo L	TGTGGCTACCTGCTGGAG	1085-1068
Homeo M	CGTGTGGTTCTTGTGCTGG	974-956
Homeo N	CACTGAGCCCCGCCAAG	3631-3647
Homeo O	AAGTTTCCTGGTGTCCCG	3869-3851
Homeo P	GGTCACGTTGCTGGAGGA	3095-3078
Homeo Q	GCCACCCCTGAAAACCA	2981-2998
Homeo R	AGTAGGGGGAGTCGTGGA	2853-2836
Homeo S	TTCCAGCAGCACAGCCCT	2404-2421
Homeo T	GCTGCACTGATGTGTGGGT	2274-2256
Homeo U	CAGACAGGAACAGCCCAG	1986-2003
Homeo B rev	GCTGATGGTGATGACGCTGA	2747-2728
Homeo D rev	ACATATTCACCCGACGCTTG	1827-1808
Homeo E rev	AGATGACCATGAAGCAGAGA	1508-1527
Homeo L rev	CTCCAGCAGGTAGCCACA	1068-1085
Homeo M rev	CCAGCACAAGAACCACACG	956-974
Homeo U rev	CTGGGCTGTTCTGTCTG	2003-1986
Homeo 2	CACCGTGATTCCCGAGAC	2279-2296

Tabela 2: Iniciadores específicos de HIPK2 utilizados no projeto

3.3 Avaliação da variação de alaninas no domínio YH em diferentes espécies

Análises comparando as seqüências de aminoácidos de Hipk2 de camundongo e PKM de hamster com a de HIPK2 humana mostraram uma variação no número de alaninas no domínio YH. Visando caracterizá-la em outras espécies e relacioná-la filogenicamente, desenhou-se iniciadores flanqueando esta região (tabela 3).

A extração do DNA das diversas espécies consistiu em primeiramente fragmentar o tecido com auxílio de uma gilete e em seguida digeri-lo em um tubo falcon de 15ml com 100mg/ml proteinase K em 10mM de Tris-HCL, pH8,0, 10mM de EDTA, pH8,0, e 0,6% SDS por 14 a 16h incubado a 37°C. À amostra foi acrescentado igual volume de fenol saturado com 1M de Tris-HCL, pH8,0, e esta foi agitada por 5 minutos e centrifugada por 10 minutos a 320 RCF a 4°C. O sobrenadante foi transferido para um novo tubo e acrescentou-se igual volume de clorofórmio/álcool isoamílico (24/1). Agitou-se novamente a amostra por 5 minutos e esta foi então centrifugada por 10 minutos a 320 RCF a 4°C. O sobrenadante foi novamente transferido para um novo tubo e tratado com clorofórmio/álcool isoamílico conforme descrito anteriormente. O sobrenadante foi transferido para tubos eppendorf de 2,0ml e o DNA precipitado com 2,5 volumes de etanol absoluto e 1:10 de volume de acetato de sódio 0,3M por 30 minutos a -70°C. A amostra foi então centrifugada a 15.300 RCF por 20 minutos e o precipitado foi lavado com etanol 70% e centrifugado por mais 10 minutos a 15.300 RCF. O precipitado foi seco e ressuspendido em água. A concentração do DNA foi determinada por leitura espectrofotométrica a 260/280nm.

Para a reação de amplificação do DNA das diferentes espécies foram realizadas diversas diluições específicas para cada animal, variando a concentração de DNA de 5 a 50µg. A única enzima capaz de amplificar estes fragmentos com diferentes combinações de iniciadores (tabela 3) foi a *Platinum®PCR SuperMix* (GibcoBRL). Para a PCR usou-se um

volume final de 10µl, contendo 2,5pmol de cada iniciador específico e 1µl do DNA diluído. A amplificação foi realizada no termociclador *GeneAmp®PCR System 9700* (PE Applied Biosystems) e a ciclagem utilizada foi 94°C por 5 minutos, 40 ciclos de 94°C por 45 segundos, 58-60°C (segundo a padronização para cada DNA) por 1 minuto, 72°C por 45 segundos e ao fim destes 40 ciclos, 3 minutos a 72°C. Após o término da reação o produto foi visualizado em gel de acrilamida 6% e a reação mantida a 4°C. A reação de sequenciamento foi a mesma já descrita no tópico 3.2e, mas antes de realizar o sequenciamento foi necessário visualizar os produtos em gel de agarose 1% corado com brometo de etídio, eluir as bandas do gel pelo *kit QIAquick™ Gel Extraction* (Quiagen) e novamente amplificá-las, obtendo assim bandas mais fortes e purificadas.

Iniciador	Seqüência 5' - 3'	Observação
Zoo 1	TCCCGCACAACAGCCCCA	
Zoo 3	ACGATGCTGGCYGGRTAGG	Y = C ou T
Zoo 4	GTGGATGGTGGGCGAGG	
Zoo 5	GCCTACATCACTCCSACC	S = G ou C
Zoo 6	GGCTGGATACTGACTSGG	S = G ou C
HomeoA	TATGTAAGGGTACTGGTTGAC	

Tabela 3: Iniciadores usados para a investigação do plimorfismo de YH.

3.4. Hibridizações em membranas de RNA total ou comerciais de RNA polyA(+)

3.4a. Preparo da sonda

A sonda para as hibridizações foi obtida a partir da digestão do clone IMAGE AA4553887 com as enzimas EcoRI e HindIII (New England). Após a digestão completa, o plasmídeo e o fragmento de interesse (usado como sonda) foram visualizados em gel de agarose 1% corado com brometo de etídio e a sonda foi purificada da agarose pelo *kit QIAquick™ Gel Extraction* (Quiagen). A sonda foi marcada com dCTP³² a partir do *kit*

Ready-to-go™ DNA labelling Beads (-dCTP) (Amersham Pharmacia Biotech). Vinte e cinco nanogramas do DNA utilizado para confecção da sonda, em um volume de 20µl, foram aquecidos a 95°C por 10min e então colocados no gelo por 1 min. Adicionou-se então 25µl de água e a mistura foi adicionada ao *bead* junto com 5µl de α -dCTP³². Após incubação a 37°C por 30 minutos, a sonda foi purificada em coluna SEPHADEX G50.

3.4b. *Preparo das membranas de RNA total*

A 15µg de RNA adicionou-se tampão de amostra numa relação volumétrica de 1:1 (este tampão deve ser feito no momento do uso e consiste em 75µl de água, 125µl de glicerol 100%, 12,5 µl de borato 50x, 1µl de EDTA 0,5M, 320µl de formamida deionizada, 100µl de formaldeído 37% e azul de bromofenol). A mistura foi aquecida a 65°C por 5 minutos e colocada em gelo por 1 minuto. Preparou-se um gel de agarose 1% tamponado com 1x Borato (1M de ácido bórico pH 8,3), 0,75 mM EDTA e 6,142% de formaldeído, que deve ser acrescentado quando a solução tiver uma temperatura menor ou igual a 55 °C. A amostra foi aplicada neste gel no tampão de corrida (1x borato, 0,5mM de EDTA e 1,85% de formaldeído) e foi realizada a eletroforese a 70V por 3h. Após a corrida, o gel foi lavado em água e foi montada a estrutura de transferência do RNA do gel para a membrana seguindo o protocolo descrito em Sambrook *et al.*, 1989. Após transferência por aproximadamente 18h, a membrana foi lavada em SSC 2x e feito um *cross-link* em UV (*Optimal cross-link – 1200J – 2x*) com a membrana ainda úmida. A integridade do RNA total e a posição do RNA 28 e 18s foram verificadas com o auxílio da luz UV e as posições foram marcadas a lápis.

3.4c. Reação de hibridização pelo método de Church & Gilbert, 1984

A membrana foi pré-hibridizada por 30 minutos a 65°C com a solução de hibridização (50% solução estoque fosfato (0,5M fosfato de sódio pH 7,2), 7% SDS, 1% BSA e 1mM EDTA). Logo após, foi acrescentada à solução de hibridização a sonda previamente fervida por 5 minutos. A hibridização foi feita a 65°C por aproximadamente 8h. A membrana então foi lavada a 65°C por 30 min com a solução I (50% solução estoque fosfato, 1% SDS e 1mM EDTA) e a seguir com a solução II também por 30 min (25% solução estoque fosfato, 1% SDS e 1mM EDTA). Foi exposto um filme para raio-X a -70°C pelo tempo necessário para visualização do resultado ou no *Storm 840*, Molecular Dynamics, pelo tempo necessário. Para remoção da sonda a membrana foi fervida duas vezes por 10 a 15 minutos (no microondas) em 20mM de Tris-HCl pH 7,5 e 0,5% SDS.

3.4.d. Reação de hibridização nas membranas da Clontech

Para a membrana de *Northern Blot, Human 12-Lane MTN[®] Blot* (Clontech), a membrana foi pré-hibridizada por 30 minutos em 5ml da solução *ExpressHyb Hybridization Solution* (Clontech) a 65°C. A sonda previamente desnaturada foi então adicionada a esta solução. A hibridização ocorreu a 65°C por 1 hora e a membrana foi então lavada com a solução I (2x SSC e 0,05% SDS) por 30 minutos, à TA e com a solução II (0,1x SSC e 0,1% SDS) por 20 minutos a 50°C. A esta membrana foi exposto um filme para raio-X a -70°C ou o *screen* do aparelho *Storm 840* (Molecular Dynamics) pelo tempo necessário para visualização do resultado. Para remoção da sonda, uma solução de 0,5% SDS foi fervida e então adicionada a membrana por 20 minutos em agitação. Para a membrana de *Dot Blot, Human Multiple Tissue Expression (MTE[™]) Array* (Clontech), a membrana foi pré-hibridizada por 30 minutos em 5ml da solução *ExpressHyb*

Hybridization Solution (Clontech) a 65°C. Foi então adicionada a esta solução a sonda, que foi previamente fervida por 5 minutos, e a hibridização ocorreu por cerca de 18h a 65°C. A membrana foi então lavada com a solução I (2x SSC e 1% SDS) por 40 minutos, a 65°C e em seguida, pelo mesmo tempo e mesma temperatura, com a solução II (0,1x SSC e 0,5% SDS). Foi exposto um filme para raio-X a -70°C pelo tempo necessário para visualização do resultado. A remoção da sonda seguiu o mesmo protocolo do descrito acima para a remoção da membrana de *Northern Blot* da Clontech.

3. 5. Hibridização *in situ*

3.5a. Escolha e obtenção da sonda

A sonda foi obtida por RT-PCR a partir de RNA total de cérebro adulto utilizando os iniciadores S e R (tabela 2), que geraram um fragmento de 450pb que codifica 149aa da proteína HIPK2. Estes 149aa possuem os 28aa finais do domínio de interação com as homeoproteínas, um sinal de localização nuclear, 60aa de 75 do sinal de retenção nuclear e 81aa dos 96 da sequência PEST. Já que alguns destes domínios se sobrepõem, a soma de cada um destes é maior que 149aa. Este fragmento foi selecionado por seu tamanho, por possuir significativo índice de identidade (90%) com a sequência nucleotídica descrita para o ortólogo de camundongo e por não apresentar similaridades significantes com outros genes ou regiões genômicas de camundongo, podendo ser considerado como específico para o gene *HIPK2*. A amplificação deste fragmento foi feita segundo as condições de PCR descritas abaixo no tópico 3.2f.

3.5b. Clonagem da sonda em pUC

Este fragmento foi clonado no vetor pUC 18 SmaI/BAP com o *kit SureClone[™] Ligation* (Amersham Pharmacia Biotech) e usado na transformação de bactérias quimiocompetentes, já descrito no tópico 3.2d.I. A seleção de bactérias contendo o inserto desejado foi feita por PCR e o plasmídeo foi purificado por *miniprep*, como já descrito no tópico 3.2d.II.

3.5c. Digestão com enzimas de restrição e clonagem direcionada em pBluescript

O plasmídeo contendo o inserto foi quantificado em gel de agarose 1% corado com brometo de etídio com o marcador de peso molecular *low mass ladder* (GibcoBRL) e digerido com as enzimas de restrição EcoRI e HindIII (New England Biolabs) na proporção de 2U por 1µg de DNA por 4 horas. O produto da digestão de aproximadamente 450pb foi visualizado em gel de agarose 1% corado com brometo de etídio e purificado do gel com o *kit QIAquick Gel Extration* (Qiagen). O resultado desta extração foi novamente quantificado em gel de agarose 1% com o marcador de peso molecular *low mass ladder* (GibcoBRL) e clonado em pBluescript. O vetor pBluescript foi previamente digerido com as mesmas enzimas de restrição e a ligação de 120ng de inserto com 250ng pBluescript foi feita com a T4 ligase (Boehringer Mannheim) segundo as instruções do fabricante.

O plasmídeo foi inserido na bactéria segundo o protocolo já descrito acima. Foi feita então uma maxi preparação por lise alcalina (*maxiprep*) pelo *kit Wizard Maxiprep* segundo instruções do mesmo.

3.5d. Síntese da sonda de RNA

O plasmídeo pBluescript foi digerido novamente com a enzima de restrição EcoRI, para torná-lo linear. Para cada 2µl do produto linearizado acrescentou-se 8µl de H₂O DEPC, 4µl

de tampão (Boehringer Mannheim), 2 µl de DTT(Boehringer Mannheim), 2 µl de dNTPs 10mM, 2µl de dUTP 35mM marcado com digoxigenina (Boehringer Mannheim), 1 µl de inibidor de RNase (Boehringer Mannheim) e 1 µl de RNA polimerase. A solução foi homogeneizada cuidadosamente e incubada por 2h a 37°C. Adicionou-se então 2 µl de DNase (Boehringer Mannheim) e incubou-se a solução por 15 minutos a 37°C. Acrescentou-se a solução de armazenamento (100µl de H₂O DEPC, 10µl de 4M LiCl e 300µl de EtOH 100%), homogeneizou-se a solução e armazenou-a a -20°C. A qualidade da sonda foi observada em gel de agarose 1% corado com brometo de etídio. Camundongas prenhas C57BL/6 foram sacrificadas e os fetos foram coletados segundo o protocolo do NIH (National Institute of Health), *NIH Guide for the Care and Use of Laboratory Animals*. Os embriões foram classificados segundo seu estágio de desenvolvimento, descrito por Theiler (1989) e os no estágio 21 (13 dias pós-coito, dpc) foram usados. Estes embriões foram fixados por 12 horas com 4% de paraformaldeído em PBS (pH 7.4), desidratados em sacarose 30% e embebidos em *Tissue-Tek* OCT. Estes embriões foram cortados com 10 µm de espessura em lâminas SuperFrost (Fisher, Pittsburgh, PA) e armazenados a -80°C.

3.5e. Hibridização dos cortes dos embriões nas lâminas

As lâminas foram descongeladas à TA. Após estarem completamente secas, foram fixadas com paraformaldeído 4% e glutaraldeído/PBS 0,2% também à TA por 10 minutos e então colocadas em um recipiente adequado. As lâminas foram lavadas à TA por 3 vezes por 5 minutos com PBT, acrescentou-se então H₂O₂/PBT por 5 minutos para clareamento das lâminas e lavou-se novamente 3 vezes por 5 minutos à TA com PBT. As lâminas foram tratadas com 10µg/ml de proteinase K em PBT por 2 minutos e lavadas inicialmente

com 2mg/ml de glicina/PBT por 30 segundos e então por 10 minutos à TA. Foram então novamente fixadas com paraformaldeído 4% PBS 0,2% à TA por 10 minutos e lavadas 3 vezes com PBT por 5 minutos à TA. As lâminas foram então colocadas duas a duas com divisórias em sacos plásticos especiais para a hibridização e acrescentou-se 5ml da solução de pré-hibridização (Formamida 50%, 5X SSC, heparina 50µg/ml, RNA de levedura 50 µg/ml e SDS 1%) previamente aquecida a 70°C e lacrou-se os sacos. Incubou-se a 70°C com agitação constante por uma hora. Foi tomado o devido cuidado para evitar bolhas nos sacos. À solução de pré-hibridização acrescentou-se 1ml da solução de hibridização e incubou-se sob agitação constante a 70°C por aproximadamente 16 horas. Para a solução de hibridização a sonda estocada foi reconstituída, ou seja, ela foi descongelada em gelo e centrifugada por 15 minutos a 14.000rpm a 4°C. O sobrenadante foi então descartado e o precipitado foi lavado com etanol 70%. O precipitado foi seco (pode ser usada centrífuga a vácuo) para então ser ressuspensionado em 40µl de H₂O DEPC. Nove microlitros desta solução foram adicionados a 1ml da solução de pré-hibridização (obtendo-se assim a solução de hibridização) e a solução foi mantida a 70°C até o momento do uso.

No dia seguinte as lâminas foram colocadas separadamente em recipiente apropriado e lavadas três vezes com a solução 1 (formamida 50%, SSC 6% e SDS 1%) por 15 minutos a 70°C em banho sob agitação constante e então lavadas 3 vezes a 65°C por 15 minutos com a solução 2 (formamida 50%, SSC 2,4%). As lâminas foram então incubadas com a solução de TBST (para cada 1000ml, 100ml de TBS 10X, 100ml de Tween20 10% e 0,4g de levimasole) e então incubadas por uma hora com soro de ovelha 10% em TBST à TA. Novamente colocou-se as lâminas em novos sacos plásticos com anticorpo anti-digoxigenina na proporção de 1:2.000 em soro de ovelha 10%/TBST e as incubou a 4°C por 16 horas.

No terceiro dia as lâminas foram removidas dos sacos, lavadas 6 vezes com TBST, 3 vezes por 5 minutos à TA e 3 vezes por 30 minutos também à TA. As lâminas foram então lavadas 3 vezes por 10 minutos à TA com NTMT (para cada litro 20ml de 5M NaCl, 50ml de 2M Tris HCl pH 9,5, 0,5ml de 2M MgCl₂, 10ml de Tween 20 10% e 0,48g de levimasole) e incubadas no agitador em recipiente embalado em papel alumínio, com a solução de reação (200ml de NTMT, 625µl NBT/dimetil formamida e 525µl BCIP) por 16-48h horas à TA e/ou até atingir a coloração de intensidade desejada. A reação foi interrompida com 1mM EDTA/PBT por 30 minutos e as lâminas foram secas com gelvitol. Estas foram fotografadas no microscópio Zeiss Jenavert.

3.6. Análise das formas de splicing do domínio de interação com as homeoproteínas (HID)

Para a análise da região inicial do domínio HID (*homeoproteins interacting domain*) que apresentava duas isoformas do gene HIPK2, realizou-se algumas RT-PCRs utilizando dois iniciadores localizados nesta região, oligonucleotídeos D e T (tabela 2).

A síntese de cDNA de tecidos ou linhagens celulares humanos ou de roedores foi feita com 2 a 4µg de RNA total ou até 500ng de RNA poly A(+). Para a reação de amplificação os cDNAs estoques das linhagens celulares e tecidos foram diluídos em uma proporção de 1:5 a 1:20 e 1µl foi usado na PCR junto com 2,5pmoles de cada iniciador específico, D (3'TGCAAGCGTCGGGTGAATA5') e T (3'GCTGCACTGATGTGTGGGT5'), 4mM de MgCl₂, 0,25mM de dNTP, 1X *PCR Buffer*, 0,5 unidades da enzima *Taq DNA Polymerase* (Pharmacia). A amplificação foi realizada no termociclador *GeneAmp®PCR System 9700* (PE Applied Biosystems, USA) e a ciclagem utilizada foi 94°C por 5 minutos, 40 ciclos de 94°C por 45 segundos, 58°C por 45 segundos e 72°C 1 minuto, segundo o tamanho do fragmento a ser amplificado, ao fim destes 40 ciclos 3 minutos a 72°C. Após o término da

reação os tubos foram mantidos a 4°C. Os produtos da reação foram visualizados em gel de acrilamida 6% corado com prata ou gel de agarose 1,5% corados com brometo de etídio.

3.7. Análises da possível terceira isoforma de HIPK2

Os produtos das RT-PCR com os iniciadores D e T, descritas acima, apresentaram uma terceira banda de tamanho maior, representando uma possível nova isoforma nesta região. Novos iniciadores, *forward* e *reverse*, foram desenhados para esta região a partir de ESTs encontradas no íntron de maior chance de conter este novo éxon. A tabela 4 lista estes novos iniciadores e a EST em que foram baseados. A RT-PCR foi feita segundo as mesmas condições citadas no tópico 3.6 e a combinação de iniciadores foi de D com TI1, TI2 ou TI3 e T com DI1, DI2 ou DI3. A reação de seqüenciamento está descrita no tópico 3.2e.

Iniciador	Seqüência 5'-3'	EST
DI1	ACATCCATACACACCCTACT	W38020
TI1	AAGCCAGAAAAAGACATCAC	W38020
DI2	AAGGCTGGAAGGGAAGTAG	BI040743
TI2	CTCAACCCTCCACATTCC	BI040743
DI3	ATCTAACATCTGCCCCATC	BE675563
TI3	TACTCCTACTTCACCCCAG	BE675563

Tabela 4: Iniciadores usados para a investigação da terceira isoforma

4. Resultados

4.1. Obtenção da sequência completa do cDNA do gene humano HIPK2 e análises da sequência de nucleotídeos e de aminoácidos

4.1a. Obtenção da sequência completa do cDNA do gene humano HIPK2

A primeira sequência obtida relativa a este gene foi uma ORESTES (Open Reading frame EST (*expressed sequence tag*), DIAS NETO et al. 2000), RC4-CT0036-160799-001-C04, proveniente de uma biblioteca de tecido tumoral de cólon gerada no Projeto Genoma do Câncer Ludwig/FAPESP. Esta EST, de apenas 121 pb, dos quais os primeiros 114pb codificavam um peptídeo de 38aa com 89% de identidade com a *homeodomain-interacting protein kinase 2 (Hipk2)* descrita em *Mus musculus*. A sequência de nucleotídeos de *Hipk2* murino foi usada para a busca de novas ESTs humanas usando o programa BLAST. Foram selecionados e adquiridos 4 clones do consórcio IMAGE (números de acesso no *GenBank*: AW044305, AI580907, H85811 e AA4553887). Os produtos do seqüenciamento foram analisados utilizando novamente o programa BLAST e apenas dois demonstraram conter os insertos esperados (números de acesso no *GenBank*: H85811 e AA4553887). Para o clone H85811, possivelmente devido a algum erro durante o processo de clonagem, ou na identificação do plasmídeo de clonagem, não foi possível o seqüenciamento com o iniciador M13F nem com o iniciador T3. Sendo assim, deste clone obtivemos aproximadamente 400pb de apenas uma das extremidades de um total de 700pb. Já no clone AA4553887, foi possível obter aproximadamente 400pb de seqüências de cada extremidade do inserto, de um total final de aproximadamente 1500pb. Para obtenção da sequência completa do cDNA de *HIPK2* foram desenhados iniciadores baseados nos resultados do seqüenciamento dos clones do consórcio IMAGE, que representavam a

região 3' deste gene e em seqüências de ESTs humanas depositadas no *GenBank* e similares a região 5' de *Hipk2* (figura 1). Em paralelo foi feita uma *Shot Gun* PCR dos clones do IMAGE.

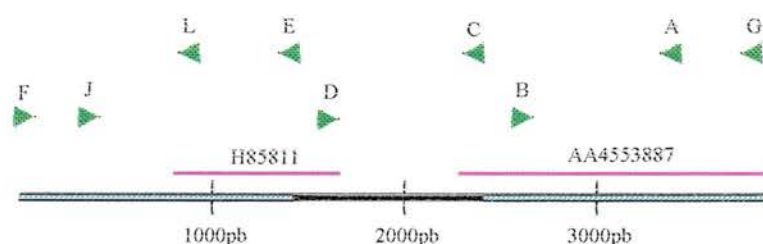


Figura 1: Esquema da obtenção da seqüência completa do cDNA de HIPK2. A linha maior representa o cDNA completo de HIPK2, em azul está representado a região que se realizou *Shot Gun* PCR, as linhas vermelhas indicam a posição dos fragmentos obtidos pelo seqüenciamento dos clones do IMAGE e as setas representam os iniciadores utilizados.

Foi obtida uma mini-biblioteca de PCR *Shot Gun* de cada clone do IMAGE, cada uma derivada da mistura de 3 reações de PCR utilizando iniciadores degenerados (figura 2). Após o seqüenciamento dos diferentes clones foi possível obter a seqüência completa do clone H85811 e quase completa do clone AA4553887, restando apenas 300pb de sua região mais 3'. Com os iniciadores A e B foi possível obter esta região mais 3' deste clone.



Figura 2: Colônias obtidas pela metodologia de *Shot Gun*. As canaletas de 1-11 são o resultado da PCR de algumas colônias obtidas na biblioteca de *Shot Gun* do clone AA4553887, mostrando a diversidade de tamanhos das bandas. M indica o marcador de peso molecular de 100pb (GibcoBRL).

Os outros pares de iniciadores, B e G, C e D e E e F, foram testados em bibliotecas de cDNAs provenientes de tecidos de cérebro fetal (Clontech, USA), tumor de mama, tumor de testículo e tumor de cólon. Mesmo alterando a temperatura de anelamento, a concentração dos iniciadores, o número de ciclos e utilizando *hot start* (Supermix High-

fidelity, GibcoBRL), não foi possível a amplificação. Foi feita uma PCR com iniciadores específicos para GAPDH, um gene *housekeeping*, para testar a qualidade das bibliotecas. Todas apresentaram a banda específica de GAPDH, indicando que a concentração utilizada para a PCR não causou uma inibição da amplificação, não sendo este o motivo da falta dessa. Paralelamente foi realizada uma hibridização com uma membrana de *Dot Blot* (Clontech), que será posteriormente comentada, onde se observou maior expressão de *HIPK2* em diversas regiões do cérebro adulto. Foi então sintetizado cDNA a partir de RNA total de lobo frontal de cérebro adulto e de glioblastoma utilizando oligo pd(T)₁₂₋₁₈ e pd(N)₆. A amplificação foi feita utilizando as três diferentes combinações de iniciadores na concentração de 2,5µM e a enzima *Supermix HI fidelity* (GibcoBRL). Apenas uma combinação não funcionou, a dupla BG, onde posteriormente o iniciador G foi substituído pelo iniciador A, mais interno (figura 1), e foi possível amplificar também este fragmento. O produto da amplificação com os iniciadores E e F, localizados na região 5' do gene resultou em uma banda de aproximadamente 1500pb no gel de agarose 1%. Com o seqüenciamento deste produto obteve-se aproximadamente 400pb em cada extremidade. Para a obtenção de sua seqüência completa foi novamente realizada a *Shot Gun* PCR e desenhados em paralelo novos iniciadores mais internos, denominados J e L (figura 1). Todas as seqüências foram agrupadas com o programa CAP3 de modo a gerar a seqüência consenso mostrada na figura 3. Esta seqüência foi depositada no *GenBank* sob o código de acesso AF326592.

cggggaggcgggtgatgcggggcgcgggcgggcctcggtgcgcgcgagagcggagacacaggetcaagatggcagattccgact 80
 gaggetggggggcgagctcgcgcgcgctttcccggtcccggttgccatgaaccgcggacaccccgcccgatggccc 160
 1 M A
 ccgtgtacgaaggtatggcctcacatgtgcaagttttctccctcacacccttcaatcaagtgcctttctgtagtgtgaag 240
 3 P V Y E G M A S H V Q V F S P H T L Q S S A F C S V K
 aaactgaaaatagagccgagttccaactgggacatgactgggtacgggtccacacagcaaagtgtatagccagagcaagaa 320
 30 K L K I E P S S N W D M T G Y G S H S K V Y S Q S K N
 catcccccgtgcgcagccagccaccacaaccgtcagcacctccttgccgggtcccaaacccaagcctaccttacgagcaga 400
 57 I P L S Q P A T T T V S T S L P V P N P S L P Y E Q
 ccacgtcttcccggaagcacgggcacatcggtgacacacagcagcactttgtcaccgggcaagtcctcggc 480
 83 T I V F P G S T G H I V V T S A S S T S V T G Q V L G
 ggaccacacaacctaagtcgctgaagcactgtgagcctccttgatacctacaaaaatgtggactcaagcgtaagagcga 560
 110 G P H N L M R S T V S L D T Y Q K C G L K R K S E
 ggagatcgagaacacaagcagcgtgcagatcatcgaggagcatcccccattgattcagaataatgcaagcggggccactg 640
 137 E I E N T S S V Q I I E E H P P M I Q N N A S G A T
 tcgccactgccaccacgtctactgccacctccaaaacagcggtcccaacagcgaggcgactatcagctgggtgcagcat 720
 163 V A T T A T T V S T S L P V P N P S L P Y E Q
 gaggtgctgtgctccatgaccaacacctacgaggtcttagagttcttgggccgagggagcgtttgggcaagtggtaagt 800
 190 E V L C S M T N T Y E V L E F L G R G T F G Q V V K C
 ctggaaacggggcaccaatgagatcgtagccatcaagatcctgaagaaccacccatcctatgccgacaaggtcagattg 880
 217 W K R I V A I K I L K N H P S Y A R Q G Q I
 aagtgcagcatcctggcccggttgagcacggagagtgcccgatgactataacttcgtccgggctacgaatgctccagcac 960
 243 E V S I L A R L S T E S A D D Y N F V R A Y E C F Q H
 aagaaccacacgtgcttggtcttcgagatgttgagcagaaacctctatgactttctgaagcaaaacaagtttagccctt 1040
 270 K N H T L N F E M L V S L G L V I A E L F L G W P L Y P
 gccctcacaatacattcgcccgagttctccagcaggtgagccacagccctgatgaaactcaaaagcctaggtcttattccacg 1120
 297 P L K Y I R P V L Q Q V A T A L M K L K S L G L I H
 ctgacctcaaacagaaaacatcatgctggtggtccatctagacaaccatacagagtcgaaggtcatcgactttggttca 1200
 323 A D L K P E N I M L V D P S R Q P Y R V K V I D F G S
 gccagccacgtctccaaggtgctgtgctccacctacttgacgtccagatattacaggggccctgagatcatccttggttt 1280
 350 A S H V S K A V C S T Y L Q S R Y Y R A P E I I L G L
 accattttgtgaggcaattgacatgtggtccctggctgtgttattgcagaattgttccctgggtggcggttatatccag 1360
 377 P F C E A I D M W S L G C V I A E L F L G W P L Y P
 gagcttcggagtgatgatcagattcggtatatttcacaaacacagggtttgctgctgaatatttattaagcgccgggaca 1440
 403 G A S E Y D Q I R Y I S Q T Q G L P A E Y L L S A G T
 aagacaactaggtttttcaacogtgacacggactcaccattcctttgtggagactgaagacaaccagatgaccatgaagc 1520
 430 K T R E R D T D S P Y P L W R L K T P D D H E A
 agagacagggattaagtcaaaagaagcaaaagtaacattttcatctgttttagatgatattggccaggtgaacatgaacga 1600
 457 E T G I K S K E A R K Y I F I C L D D M A Q V N M T
 cagatttgaagggagcgacatgttggttagaaaaggtgacccggcgggagttcattgacctgttgaagaagatgctgacc 1680
 483 T D L E G S D M L V E K A D R E F I D L L K M L T
 attgatgctgacaagagaatcactccaatcgaaacctgaacctccctttgtccacatgacacacttactcagattttcc 1760
 510 I D A D K R I T P I E T L N H P F V T M T H L L D F P
 ccacagcacacacgtcaaatcatgtttccagaacatggagatctgcgaagcgtcggtgaatatgtatgacacgggtgaacc 1840
 537 H S T H V K S C F N M E I C K R R V N M Y D T V P
 agagcaaaacccctttcatcagcagcgtggccccagcagcgtccaccaacctgaccttaaacaccagctgacc 1920
 563 Q S K T P F I T H V A P S T S T N L T M T F N N Q L T
 actgtccacaaccagccctcagcggcatccatggctgcagtgggccagcggagcagccttcagacaggaacagccca 2000
 590 T V H N Q P S A A S M A V A Q R S M P L Q T G A Q
 gatttgtgcggcggtgaccoggtccagcaagctctcatctgtgtgtcccccgggttccaaggttgaggcctctccct 2080
 617 I C A R P D P E Q Q A L I V C P P G F Q G L Q A S P
 ctaagcagcgtggtactcgggtgcgaatggaaaatgcagttcccatcgtcactcaagccccaggagctcagcctcttcag 2160
 643 S K H A G Y S V R M E N A V P I V T Q A P G A Q P L Q
 atccaaccaggtctcttgccagcaggtcttgcccaagtgaggagccagcagatcctgctcccccagcatggcgacaa 2240
 670 I Q P G L L A Q Q A W P S G T Q Q I L L P P A W Q Q L
 gactggagtgccacccacacatcagtgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2320
 697 T G V A T H T S V Q H A T V I P E T M A G T Q L A
 actgggaaaatacgcagc 2400
 723 D W R N T H A H G S H Y N P I M Q Q P A L L T G H V T
 cttccagc 2480
 750 L P A A Q P L N V G V A H V M R Q Q P T S T S R K
 gagtaagc 2560
 777 S K Q H Q S S V R N V S T C E V S S S Q A I S S P Q
 gatccaagcgtgtcaaggagaaacacacctcccgctgtgccatgggtgcacagtagcccggtcgtcagcagcctcggtcacc 2640
 803 R S K R V K E N T P P R C A M V H S S P A C S T S V T
 tgtgggtggggcgagc 2720
 830 C G W G D V A S S T T R E R Q R Q T I V I P D T P S P
 cagcgtcagcgtcatcaccatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2800
 857 T V S V I T I S D T D E E E Q K H A P T S T V S
 agcaaaagaaacagc 2880
 883 K Q R K N V I S C V T V H D S P Y S D S S S N T S P Y
 tccgtgcagc 2960
 910 S V Q Q R A G H N N A N A F D T K G S L E N H C T G N
 ccccggaaccatcatcgtgcccacctgaaaacccagggcagcagcagcagcagcagcagcagcagcagcagcagcagc 3040
 937 P R T I I V P P L K T Q A S E V L V E C D S L V P V
 acaccagtcaccactcgtcctcctacaagtccaagtcctccagcaacgtgacctccaccagcggctcactcttcagggagc 3120

```

963 N T S H H S S S Y K S K S S S N V T S T S G H S S G S
    tcattctggagccatcacctaccggcagcagcgccgggcccccaattccagcagcagcagccactcaatctcagccaggc 3200
990 S G A I T Y R Q Q R P G P H F Q Q Q Q P L N L S Q A Q
    tcagcagcacatcaccacggaccgcactgggagccacgaaggcagcagcgctacatcactcccaccatggcccaggctc 3280
1017 Q H I T T D R T G S H R R Q Q A Y I T P T M A Q A P
    cgtactccttcccgacacagccccagccacggcactgtgcaccgcgcatctggctgcagccgctgcccgtgcccacctc 3360
1043 Y S F P H N S P S H G T V H P H L A A A A A A A H L P
    cccacccagccccacctctacacctacactgcccggcgccgctgggctccaccggcaccgtggcccacctgggtggcctc 3440
1070 T Q P H L Y T Y T A P A A L G S T G T V A H H L V A S
    gcaaggctctgcccggccacacgtgcagcacactgcctaccagccagcatcgtccaccagggtcccgtgagcatgggccc 3520
1093 Q G S A R H T V Q H T A Y P A S I V H Q V P V S M G
    cccgggtcctgcctcgcgccaccatccacccgagtcagtatccagcccaatttgcaccaccagacctacatcagcgccctc 3600
1123 P R V L P S D R T I H P S Q Y P A Q F A H Q T Y I S A S
    ccagcctccaccgctctacactggataccactgagccccgccaagggtcaaccagtagcccttacatataaacactggagg 3680
1150 P A S T V Y T G Y P L S P A K V N Q Y P Y I
    gagggaggggaggggaggaggagagaatggcccgaggaggaggagagaaggaggaggcgctcctgggaccgtggggcg 3760
    ctggccttttatactgaagatgccgcacacaaacaatgcaaacggggcagggcgggggggggggggggggggggggggg 3840
    ggacgggtcgggacaccagtgaacttgaaccgggaagtgggagacgtagagcagagaagagaacatttttaaaaggaa 3920
    gggattaaagggggtgggaaatctatggtttttattttaaaaaaaaaaaaaaaaaaaaaa 3982

```

Figura 3: Seqüência completa de nucleotídeos e a seqüência deduzida de aminoácidos (aa) do cDNA de HIPK2. Em letras minúsculas está representada seqüência de nucleotídeos de HIPK2 obtida por RT-PCR e em letras maiúsculas a seqüência de aa deduzida pelos programas *GeneRunner* e *BlastX*. Em azul destacam-se os dois códons de iniciação e o de término da proteína e em rosa o sinal de poliadenilação. Os números à esquerda indicam o número de aa e os da direita o número de nucleotídeos.

4.1b. Domínios da Proteína Codificada

Utilizando os programas *GeneRunner* e *BlastX* obtivemos a possível seqüência de aa de HIPK2 (figura 3) que apresentou 94% de identidade com Hipk2 (camundongo) e 96% com a proteína PKM (*Mx-interacting protein kinase*) de hamster. Para as nossas análises comparativas de domínios nos basearemos na proteína de camundongo. Na HIPK2 humana encontramos os seguintes domínios (figura 5):

1) Domínio serina-treonina quinase com homologia com outras proteínas quinases de diferentes espécies (figura 4), como com a proteína hPKY humana (87% de identidade), o produto gênico CG01790 de drosófila (83% de identidade) e a proteína hipotética F2B6.8 de *C. elegans* (69% de identidade);

2) Domínio HID (*homeoproteins-interacting domain*) que, diferente do de Hipk2, não possui os primeiros 27aa e tem uma identidade restante de 97% com o domínio da proteína do camundongo. A proteína PKM também não possui estes 27aa;

3) Domínio co-repressor com 100% de identidade com o de Hipk2;

4) Dois sinais de localização nuclear (NLS, *nuclear localization signal*) tendo o primeiro 100% e o segundo 98% de identidade com a proteína murina;

5) Um sinal de retenção no *nuclear speckle* (SRS) com 94% de identidade com o encontrado em Hipk2;

6) A sequência PEST, prolina (P), ácido glutâmico (E), serina (S) e treonina (T), com 100% de identidade;

7) O domínio YH, uma região enriquecida com resíduos de histidina e tirosina que apresenta 95% identidade. Nesta região foi encontrada uma variação no número de alaninas entre o camundongo e o humano. A proteína humana possui sete alaninas nesta região enquanto a proteína do rato possui apenas quatro (figura 5). Uma avaliação da variação desta região em diferentes espécies é apresentada no tópico 4.1d.

HIPK2	1	EVLE LGRG T GQVVKC K RGTNEIVAIAK ILKNHPS AR QGQIEVSILA
Hipk2	1	EVLE LGRG T GQVVKC K RGTNEIVAIAK ILKNHPS AR QGQIEVSILA
PKM	1	EVLE LGRG T GQVVKC K RGTNEIVAIAK ILKNHPS AR QGQIEVSILA
hPKY	1	EVLD LGRG T GQVVKC K RGTNEIVAIAK ILKNHPS AR QGQIEVSILA
Hipk1	1	EVLE LGRG T GQVAKC K RSTKEIVAIAK ILKNHPS AR QGQIEVSILS
HIPK3	1	EVLD LGRG T GQVVKC K RGTNEIVAIAK ILKNHPS AR QGQIEVSILA
dCG17090	1	EVLE LGRG T GQVVKC K RGTSEIVAIAK ILKNHPS AR QGQIEVSILS
cF2B6.8	1	EVLE LGKG T GQVVKC K KGTSEIVAIAK ILKKHPS AR QGQIEVSILS
yYak1p	1	LVLDILGQG T GQVVKCQN LLTKEILAVK VVKSRTTE LT QSITEAKILE
HIPK2	51	RLSTESA--D D N VRA EC QHKNHHTCLV EMLEQNL D LKQNK SPL
Hipk2	51	RLSTESA--D D N VRA EC QHKNHHTCLV EMLEQNL D LKQNK SPL
PKM	51	RLSTESA--D D N VRA EC QHKNHHTCLV EMLEQNL D LKQNK SPL
hPKY	51	RLSTENA--D E N VRA EC QHRNHHTCLV EMLEQNL D LKQNK SPL
Hipk1	51	RLSSENA--D E N VRS EC QHKNHHTCLV EMLEQNL D LKQNK SPL
HIPK3	51	RLSTENA--D E N VRA EC QHRNHHTCLV EMLEQNL D LKQNK SPL
dCG17090	51	RLSQENA--D E N VRA EC QHKNHHTCLV EMLEQNL D LKQNK SPL
cF2B6.8	51	RLSNENS--E E N VRA EC NHKSHTCLV EMLEQNL D LKQNK MPL
yYak1p	51	LLNQKidptN KHH LRM DS VHKNHLCV ELLSNNL E LLKQNK hqL
HIPK2	99	PLK IRPVLQ QVATALMKLK SLGLIHADLK PENIMLVDP S RQP RVKVID
Hipk2	99	PLK IRPVLQ QVATALMKLK SLGLIHADLK PENIMLVDP S RQP RVKVID
PKM	99	PLK IRPVLQ QVATALMKLK SLGLIHADLK PENIMLVDP S RQP RVKVID
hPKY	99	PLKVIRPILQ QVATALKKLK SLGLIHADLK PENIMLVDPV RQP RVKVID
Hipk1	99	PLK IRPILQ QVATALMKLK SLGLIHADLK PENIMLVDPV RQP RVKVID
HIPK3	99	PLKVIRPILQ QVATALKKLK SLGLIHADLK PENIMLVDPV RQP RVKVID
dCG17090	99	PLK IRPILE QVLTALLKLK QLGLIHADLK PENIMLVDPV RQP RVKVID
cF2B6.8	99	PLNAIRPIL QVLTALLKLK SLGLIHADLK PENIMLVDPQ QQP RVKVID
yYak1p	101	SIQLIRT TT QILDSLCLVK ESKLIHCDLK PENILLCAPd kpe--LKIID
HIPK2	149	GSASHVSKA VCST LQSR RAPEIILGL P CEAIDM S LGCVIAEL L
Hipk2	149	GSASHVSKA VCST LQSR RAPEIILGL P CEAIDM S LGCVIAEL L
PKM	149	GSASHVSKA VCST LQSR RAPEIILGL P CEAIDM S LGCVIAEL L
hPKY	149	GSASHVSKT VCST LQSR RAPEIILGL P CEAIDM S LGCVIAEL L
Hipk1	149	GSASHVSKA VCST LQSR RAPEIILGL P CEAIDM S LGCVIAEL L
HIPK3	149	GSASHVSKT VCST LQSR RAPEIILGL P CEAIDM S LGCVIAEL L
dCG17090	149	GSASHVSKT VCNT LQSR RAPEIILGL P CEAIDM S LGCVVAEL L
cF2B6.8	149	GSASHRSKA VtNT LQSR RAPEIILGL P NESIDM S LGCVIAEL L
yYak1p	149	GSSCEEART Vy-T IQSR RAPEIILGI P STSIDM S LGCVIAEL L

HIPK2	199	G PL PGASE	DQIR ISQT	QGLPAE LLS	AGTKTTR	N R-----DT
Hipk2	199	G PL PGASE	DQIR ISQT	QGLPAE LLS	AGTKTTR	N R-----DT
PKM	199	G PL PGASE	DQIR ISQT	QGLPAE LLS	AGTKTTR	N R-----DT
hPKY	199	G PL PGASE	DQIR ISQT	QGLPGEQLLN	VGTKSTR	C K-----ET
Hipk1	199	G PL PGASE	DQIR ISQT	QGLPAE LLS	AGTKTTR	N R-----D
HIPK3	199	G PL PGASE	DQIR ISQT	QGLPGEQLLN	VGTKSTR	C K-----ET
dCG17090	199	G PL PGSSE	DQIR ISQT	QGLPTEHMLN	SASKTSK	R-----DV
cF2B6.8	199	G PL PGSSE	DQIR IIQT	QGLPPTSMLE	SASKLHR	K EVkseSPNHT
yYak1p	198	GiPI PGASE	NQLTRIIDT	LG PPS MID	MGKNSGK	MK KL---APEES
HIPK2	242	DSP PL RLK	TPDDHEAETG	I-KSKEARK	I ICL--DDM	AQVNMTTDLE
Hipk2	242	DSP PL RLK	TPDDHEAETG	T-KSKEARK	I NCL--DDM	AQGNMTTDLE
PKM	242	DSP PL RLK	TPDDHEAETG	I-KSKEARK	I NCL--DDM	AQVNMTTDLE
hPKY	242	DMSHSG RLK	TLEEHEAETG	M-KSKEARK	I NSL--DDV	AHVNTVMDLE
Hipk1	242	NLG PL RLK	TPEEHELETG	I-KSKEARK	I NCL--DDM	AQVNMSTDLE
HIPK3	242	DMSHSG RLK	TLEEHEAETG	M-KSKEARK	I NSL--DDV	AHVNTVMDLE
dCG17090	242	DST P RLK	TTEEHEAETN	T-KSKEARK	I NCL--DDI	GQVNVPTDLE
cF2B6.8	249	NVGGGS RLK	TVEE EASSS	TaKSKETRK	I NVI--DDI	SRVcygfesd
yYak1p	245	SSSTQKHRMK	TIEE CRE N	I-VEKPSKQ	kwrk1pDII	RN R PKSIQ
HIPK2	289	GSDMLV--EK	ADRRE IDLL	KKMLTIDADK	RITPIETLNH	P V-
Hipk2	289	GSDMLV--EK	ADRRE IDLL	KKMLTIDADK	RVTPiETLNH	P V-
PKM	289	GSDMLV--EK	ADRRE IDLL	KKMLTIDADK	RITPIETLNH	P V-
hPKY	289	GSDLLA--EK	ADRRE VSLL	KKMLLIDADL	RITPAETLNH	P V-
Hipk1	289	GTDLMA--EK	ADRRE IDLL	KKMLTIDADK	RITPLKTLNH	Q V-
HIPK3	289	GSDLLA--EK	ADRRE VSLL	KKMLLIDADL	RITPAETLNH	P Vn
dCG17090	289	GGQLLA--EK	TDRRE IDLL	KRMLTIDQER	RLTPAEALNH	S --
cF2B6.8	297	pvehLC--DR	IDRQE VDVL	KKMLVLNPD	RITPAEGLES	K V-
yYak1p	294	NSQELIdqEM	QNRECLIH L	GGVLNLNPLE	R TPQQAMLH	P I-

Hipk2	-----MASHVQVFSPHTLQSSAFCSVKKLKVEPSSNWDMTGYGSHSKVYSQSKNIPPS	53
HIPK2	MAPVYEGMASHVQVFSPHTLQSSAFCSVKKLKIEPSSNWDMTGYGSHSKVYSQSKNIPLS	60
	*****:***** *	
Hipk2	QPASTTVVSTSLPIPNPSLPYEQTIIIFPGSTGHIVVTSASSTSVTGQVLGGPHNL MRRSTV	113
HIPK2	QPATTTVVSTSLFPVNPSPLPYEQTIVFPGSTGHIVVTSASSTSVTGQVLGGPHNL MRRSTV	120
	:**:*****:*****	
Hipk2	<u>SLLDTYQKCGLKRKSEEIENTSSVQIIIEHPPMIQNNASGATVATATTSTATSKNSGSNS</u>	173
HIPK2	<u>SLLDTYQKCGLKRKSEEIENTSSVQIIIEHPPMIQNNASGATVATATTSTATSKNSGSNS</u>	180

Hipk2	<u>EGDYQLVQHEVLCSMTNTYEVLEFLGRGTFGQVVKCWKRGTNEIVAIKILKNHPSYARQG</u>	233
HIPK2	<u>EGDYQLVQHEVLCSMTNTYEVLEFLGRGTFGQVVKCWKRGTNEIVAIKILKNHPSYARQG</u>	240

Hipk2	<u>QIEVSILARLSTESADDYNFVRAYECFQHKNHTCLVFEMLEQNLNLYDFLKQNKFSPLPLKY</u>	293
HIPK2	<u>QIEVSILARLSTESADDYNFVRAYECFQHKNHTCLVFEMLEQNLNLYDFLKQNKFSPLPLKY</u>	300

Hipk2	IRPVLQQVATALMKLKSGLIHADLKPENIMLVDPSPQPYRVKVIDFGSASHVSKAVCST	353
HIPK2	IRPVLQQVATALMKLKSGLIHADLKPENIMLVDPSPQPYRVKVIDFGSASHVSKAVCST	360

Hipk2	YLQSRYYRAPEIILGLPFCEAIDMWSLGCVIAELFLGWPLYPGASEYDQIRYISQTQGLP	413
HIPK2	YLQSRYYRAPEIILGLPFCEAIDMWSLGCVIAELFLGWPLYPGASEYDQIRYISQTQGLP	420

Hipk2	AEYLLSAGTKTTRFFNRDTSYPYPLWRLKTPDDHEAETGTSKEARKYIFNCLDDMAQGN	473
HIPK2	AEYLLSAGTKTTRFFNRDTSYPYPLWRLKTPDDHEAETGIKSKEARKYIFICLDDMAQVN	480

Hipk2	MTDLEGSMDLVEKADRREFIDLKKMLTIDADKRVTPIETLNHPFVTMTHLLDFPHSAH	533
HIPK2	MTDLEGSMDLVEKADRREFIDLKKMLTIDADKRITPIETLNHPFVTMTHLLDFPHSTH	540

Hipk2	VKSCFQNMIEICKRRVNMVDTVNQSKTPFITHVAPSTSTNLMTENNQLTTVHNQAPTSS	593
HIPK2	VKSCFQNMIEICKRRVNMVDTVNQSKTPFITHVAPSTSTNLMTENNQLTTVHN-----	593

Hipk2	ATLSLANPEVSILNYQSALYQPSAASMAAVAPRSMPLQTGTAQICARPDFQQAALIVCPP	653
HIPK2	-----QPSAASMAAVAQPSMPLQTGTAQICARPDFQQAALIVCPP	633

Hipk2	GFQGLQASPSKHAGYSVRMENAVPIVTQAPGAQPLQIQPGLLAQQAWPGGAQQIILLPPTW	713
HIPK2	GFQGLQASPSKHAGYSVRMENAVPIVTQAPGAQPLQIQPGLLAQQAWPSGTQQIILLPPAW	693

Hipk2	QQLTGVAHTSVQHAIVI PETMAGTQQLADWRNTHAHGSHYNPIMQQPALLTGHVTLPA	773
HIPK2	QQLTGVAHTSVQHATVI PETMAGTQQLADWRNTHAHGSHYNPIMQQPALLTGHVTLPA	753

Hipk2	QPLNVGVAVHMRQQPTSTTSRKSQHQSSVRNVSTCEVTSQAISSPQSRKRVKENTPP	833
HIPK2	QPLNVGVAVHMRQQPTSTTSRKSQHQSSVRNVSTCEVSSQAISSPQSRKRVKENTPP	813

Hipk2	RCAMVHS SPACSTSVTCGWGDVASST TRERQRQTIVIPDTPSPTVSVITISSDTDEEEEQ	893
HIPK2	RCAMVHS SPACSTSVTCGWGDVASST TRERQRQTIVIPDTPSPTVSVITISSDTDEEEEQ	873

Hipk2	<u>KHAPTSTVSKQRKNVISCVTVHDSPYSDSSNTSPYSVQRTGHNGTNTLDTKGLENHC</u>	953
HIPK2	<u>KHAPTSTVSKQRKNVISCVTVHDSPYSDSSNTSPYSVQRTAGHNNANAFDTKGLENHC</u>	933

Hipk2	<u>TGNPRTIIVPPLKTQASEVLVECDSLGPAISASHHSSSFKSSSTVTSTSGHSSGSSSG</u>	1013
HIPK2	<u>TGNPRTIIVPPLKTQASEVLVECDSLVP-VNTSHHSSSYKSSSNVTSTSGHSSGSSSG</u>	992

Hipk2	AIAYRQQRPGPHFQQQQLNLSQAQQAADRTGSHRRQQAYITPTMAQAPYTFPHNSPS	1073
HIPK2	AITYRQQRPGPHFQQQQLNLSQAQQAADRTGSHRRQQAYITPTMAQAPYSFPHNSPS	1052

Hipk2	HGTVHPHL AAAA ---HLPTQPHLYTYTAPTALGSTGTVAHLVRSQGSARHTVQHTAYPAS	1130
HIPK2	HGTVHPHL AAAAAA HLPTQPHLYTYTAPALGSTGTVAHLVASQGSARHTVQHTAYPAS	1112

Hipk2	IVHQVPVSMGPRVLPSTIHPSQYPAQFAHQTYISASPASTVYTGYPSPAKVNQYPYI	1189
HIPK2	IVHQVPVSMGPRVLPSTIHPSQYPAQFAHQTYISASPASTVYTGYPSPAKVNQYPYI	1171

Figura 5: Comparação entre a seqüência de aa de Hipk2 e HIPK2. A proteína HIPK2 possui 94% de identidade com Hipk2. O domínio co-repressor está sublinhado, os sinais de localização nuclear estão em negrito, o domínio quinásico está representado em vermelho, o domínio de interação com as homeoproteínas está em verde, a seqüência PEST está em azul, o sinal de retenção no *nuclear speckle* está em itálico e sublinhado, o domínio YH está em roxo e em negrito e em rosa e itálico está apontada a região de poli-alaninas polimórfica entre a proteína humana e a de camundongo.

4.1c. Análises Genômicas

Análises de BlastN da sequência de cDNA de *HIPK2* mostraram a existência de 3 clones de DNA genômico, o PAC clone RP5-1129624 de *Homo sapiens* (número de acesso no *GenBank*: AC006021) com localização entre 7q32 e 7q34, com homologia entre o primeiro par de bases do gene a 177pb, o PAC clone RP4-701016 de *Homo sapiens* (número de acesso no *GenBank*: AC005531) com localização entre 7q33 e 7q36, com homologia a partir de 2792pb até o final do gene e o clone CTD-2022G18 de *Homo sapiens* (número de acesso no *GenBank*: AC073184) com homologia entre 1259pb e 2791pb. A figura 6 e a tabela 5 fazem uma relação entre os éxons e íntrons encontrados nas análises de bancos de dados de sequências genômicas públicas a partir da posição 1259pb de *HIPK2*.

Devido à compra do acesso ao banco de sequências da Celera feita pelo Instituto Ludwig, foi possível complementar nossos dados. Neste banco nosso gene de interesse foi encontrado quase por completo, mas visto que ainda não houve o agrupamento total das sequências a região entre 177pb e 1259pb não estava ainda mapeada. Porém no banco da Celera um clone não mapeado de 685pb com homologia na sua extremidade final de maneira decrescente com o nucleotídeo 863 até 1259 de *HIPK2*. A figura 7 ilustra um possível mapeamento de *HIPK2*. Análise da sequência genômica de *HIPK2* pelo programa LocusLink (<http://www.ncbi.nlm.nih.gov/LocusLink/list.cgi>) indica que este gene está localizado no cromossomo 7q32-34.

Início do éxon (pb)	Tamanho do éxon (pb)	Tamanho do intron (pb)
1259	124	?
1383	120	317
1503	87	2.125
1590	185	2.167
1775	163	6.037
1938	127	5.998
2065	122	962
2187	143	8.978
2330	180	3.484
2510	282	3.418
2792	249	12.652
3040	161	8.528
3201	769	1.730

Tabela 5: Organização parcial de éxons e introns deste gene.

aggagtcagt	aatggcacc	caggtttgaa	agatagtatt	catatagaaa	
cagcttcctg	catttccaac	ttgctaattg	tttcttccat	tcttcctttt	
cagggcccc	GAGATCATCC	TTGGTTTACC	ATTTTGTGAG	GCAATTGACA	
TGTGGTCCCT	GGGCTGTGTT	ATTGCAGAAT	TGTTCTGCGG	TTGGCCGTTA	
TATCCAGGAG	CTTCGGAGTA	TGATCAGgtg	ggtccagatg		317pb
atcttttctt	gcagATTTCGG	TATATTTTAC	AAACACAGGG	TTTGCCGTGCT	
GAATATTTAT	TAAGCGCCGG	GACAAAGACA	ACTAGGTTTT	TCAACCGTGA	
CACGGACTCA	CCATATCCTT	TGTGGAGACT	GAAGgtagga	agagcgtctc	2.125pb
.....	cttttgctc	tcttgacaga	CACCAGATGA	CCATGAAGCA	
GAGACAGGGA	TTAAGTCAAA	AGAAGCAAGA	AAGTACATTT	TCAaCTGTTT	
AGATGATATG	GCCCAGgtaa	gcttcatggg		tacttgctgc	2.167pb
caggtGAACA	TGACGACAGA	TTTGAAGGG	AGCGACATGT	TGGTAGAAAA	
GGCTGACCGG	CGGGAGTTCA	TTGACCTGTT	GAAGAAGATG	CTGACCATTG	
ATGCTGACAA	GAGAATCACT	CCAATCGAAA	CCCTGAACCA	TCCCTTTGTC	
ACCATGACAC	ACTTACTCGA	TTTTCCCCAC	AGCACACagt	aggtgtcctg	6.037pb
.....	cttttgcatg	aacagCGTCA	AATCATGTTT	CCAGAACATG	
GAGATCTGCA	AGCGTCGGGT	GAATATGTAT	GACACGGTGA	ACCAGAGCAA	
AACCCCTTTC	ATCACGCACG	TGGCCCCCAG	CACGTCCACC	AACCTGACCA	
TGACCTTTAA	CAACCAGCTG	ACCACTGTCC	ACAACCAGgt	aagtgggaaa	5.998pb
.....	atctacactc	taccagCCCT	CAGCGGCATC	CATGGCTGCA	
GTGGCCAGC	GGAGCATGCC	CCTGCAGACA	GGAACAGCCC	AGATTGTGTC	
CCGGCCTGAC	CCGTTCCAGC	AAGCTCTCAT	CGTGTGTCCC	CCCGGCTTCC	
AAGgttaagat	tgagcagcca		cactgtctgg	aacagGCTTG	962pb
CAGGCCTCTC	CCTCTAAGCA	CGCTGGCTAC	TCGGTGCGAA	TGGAAAATGC	
AGTTCCCATC	GTCACCTAAG	CCCCAGGAGC	TCAGCCTCTT	CAGATCCAAC	
CAGGTCTGCT	TGCCCAGgtta	attctttctc		tcctgtgttc	8.978pb
tacagCAGGC	TTGGCCAAGT	GGGACCCAGC	AGATCCTGCT	TCCCCAGCA	
TGGCAGCAAC	TGACTGGAGT	GGCCACCCAC	ACATCAGTGC	AGCATGCCAC	
CGTGATTCCC	GAGACCATGG	CAGGCACCCA	GCAGCTGGCG	GACTGGAGgt	
aagcaggaaa		gttgccctttc	agAAATACGC	ATGCTCACGG	3.484pb
AAGCCATTAT	AATCCCATCA	TGCAGCAGCC	TGCACTATTG	ACCGGTCATG	
TGACCCTTCC	AGCAGCACAG	CCCTTAAATG	TGGGTGTGGC	CCACGTGATG	
CGGCAGCAGC	CAACCAGCAC	CACCTCCTCC	CGGAAGAGTA	AGCAGCACCA	
GTCATCTGTG	AGgtlaggtgg	tgagagccac		gtgtgtgtttt	3.418pb
ctcattacag	AAATGTCTCC	ACCTGTGAGG	TGTCCTCCTC	TCAGGCCATC	
AGCTCCCCAC	AGCGATCCAA	GCGTGTCAAG	GAGAACACAC	CTCCCCGCTG	

```

TGCCATGGTG CACAGTAGCC CGGCTGCAG CACCTCGGTC ACCTGTGGGT
GGGGCGACGT GGCCTCCAGC ACCACCCGGG AACGGCAGCG GCAGACAATT
GTCATTCCCG ACACTCCCAG CCCACGGTC AGCGTCATCA CCATCAGCAG
TGACACGGAC GAGGAGGAGG AACAGAAACA CGCCCCACC AGgtgaggca 12.652pb
gtgccacctt ..... ctttgcattt tcagCACTGT CTCCAAGCAA
AGAAAAACG TCATCAGCTG TGTCACAGTC CACGACTCCC CCTACTCCGA
CTCCTCCAGC AACACCAGCC CCTACTCCGT GCAGCAGCGT GCTGGGCACA
ACAATGCCAA TGCCTTTGAC ACCAAGGGGA GCCTGGAGAA TCACTGCACG
GGGAACCCCC GAACCATCAT CGTGCCACCC CTGAAAACCC AGGCCAGCGA
AGTATTGGTG GAGTGTGATA GCCTGGTGCC AGgttaatttg gggccaggac 8.528pb
..... tctttctcct cccgccacag TCAACACCAG TCACCACTCG
TCCTCCTACA AGTCCAAGTC CTCCAGCAAC GTGACCTCCA CCAGCGGTCA
CTCTTCAGGG AGCTCATCTG GAGCCATCAC CTACCGGCAG CAGCGGCCGG
GCCCCACTT CCAGCAGCAG CAGCCACTCA ATCTCAGCCA Ggttaagcccc 1.730pb
acagctggcg ..... ctctctccca ggGCTCAGCAG CACATCACCA
CGGACCGCAC TGGGAGCCAC CGAAGGCAGC AGGCCTACAT CACTCCCACC
ATGGCCCAGG CTCCGTACTC CTTCCCGCAC AACAGCCCCA GCCACGGCAC
TGTGCACCCG CATCTGGCTG CAGCCGCTGC CGCTGCCCAC CTCCCCACCC
AGCCCCACCT TACACCTAC ACTGCGCCGG CGGCCCTGGG CTCCACCGGC
ACCGTGGCCC ACCTGGTGCC CTCGCAAGGC TCTGCGCGCC ACACCGTGCA
GCACACTGCC TACCCAGCCA GCATCGTCCA CCAGGTCCCC GTGAGCATGG
GCCCCCGGGT CCTGCCCTCG CCCACCATCC ACCCGAGTCA GTATCCAGCC
CAATTTGCC ACCAGACCTA CATCAGCGCC TCGCCAGCCT CCACCGTCTA
CACTGGATAC CCACTGAGCC CCGCCAAGGT CAACCAGTAC CCTTACATAT
AAACACTGGA GGGGAGGGAG GGAGGGAGGG AGGGAGAGAA TGGCCCGAGG
GAGGAGGGAG AGAAGGAGGG AGGCGCTCCT GGGACCGTGG GCGCTGGCCT
TTTATACTGA AGATGCCGCA CACAAACAAT GCAAACGGGG CAGGGGCGGG
GGGGGGGGGG GGGGCAGAGG GCAGGGGGGAC GGGTCGGGAC ACCAGTGAAA
CTTGAACCGG GAAGTGGGAG GACGTAGAGC AGAGAAGAGA ACATTTTTAA
AAGGAAGGGA TTAAAGAGGG TGGGAAATCT ATGGTTTTTA TTTTAAAAAA

```

Figura 6: Sequência genômica a partir da região de 1.259pb do cDNA de HIPK2 encontrada em bancos públicos. Em azul estão representados os éxons e em rosa estão indicados os sítios *donnors* e *acceptors* de *splicing*. A sequência dos introns está parcialmente representada e os números indicam o tamanho de cada intron.

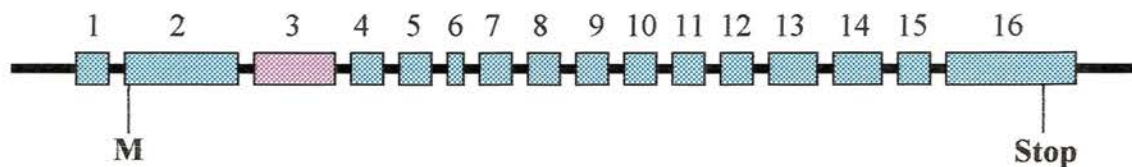


Figura 7: Desenho esquemático da distribuição de introns e éxons de HIPK2. Os quadrados indicam os éxons, sendo o quadrado rosa o provável éxon ainda não confirmado. A linha preta representa os introns. M indica a primeira metionina e Stop representa o códon de parada.

Em dezembro de 2000 foi depositada no *GenBank* por Pierantoni e colaboradores (número de acesso: AF207702) uma isoforma menor de *HIPK2* que a aqui descrita com 911 aa e 94% de similaridade com a proteína aqui estudada. Em março de 2001 Wang e colaboradores depositaram no *GenBank* uma segunda isoforma (número de acesso: AF208291) com 1.198 aa e 97% de similaridade com a nossa isoforma. Esta última diferia da nossa em apenas 27 aa no início do domínio HID, como o primeiro cDNA descrito em camundongos. A isoforma descrita em dezembro também apresentava estes 27aa. A figura 8 mostra o alinhamento destas 3 novas isoformas. Análises genômicas desta região mostram a existência de dois possíveis sítios *acceptors* de *splicing* com 81nt de distância entre eles, gerando assim as 2 isoformas (figura 9). Analisamos as ESTs mapeadas para esta região e encontramos apenas 5, das quais 4 (números de acesso no *GenBank*: BF904758, BF935452, BF814194, BF814195) não possuem os 81nt e eram ORESTES. A única que possui os 81nt foi a de número de acesso N43800 derivada de uma biblioteca de melanócito. As três únicas ESTs de camundongos nesta região possuem os 81nt (número de acesso: AA549549, AI503712, BG247714).

AF326592 MAPVYEGMAHSHVQVESPHTLQSSAFCSVKKLKIEPSSNWDMTGYGSHSKVYSQSKNIPLS 60
 AF208291 MAPVYEGMAHSHVQVESPHTLQSSAFCSVKKLKIEPSSNWDMTGYGSHSKVYSQSKNIPLS 60
 AF207702 -----MASHVQVESPHTLQSSAFCSVKKLKIEPSSNWDMTGYGSHSKVYSQSKNIPLS 53
 *****:*****

AF326592 QPATTTVSTSLPVPNPSPLEQTIIVFPGSTGHIVVTSASSTSVTGQVLGGPHNLMRRSTV 120
 AF208291 QPATTTVSTSLPVPNPSPLEQTIIVFPGSTGHIVVTSASSTSVTGQVLGGPHNLMRRSTV 120
 AF207702 QPATTTVSTSLPVPNPSPLEQTIIVFPGSTGHIVVTSASSTSVTGQVLGGPHNLMRRSTV 113
 :**

AF326592 SLDDTYQKCGLKRKSEEIENTSSVQIIEEHPPMIQNNASGATVATATTSTATSKNSGSNS 180
 AF208291 SLDDTYQKCGLKRKSEEIENTSSVQIIEEHPPMIQNNASGATVATATTSTATSKNSGSNS 180
 AF207702 SLDDTYQKCGLKRKSEEIENTSSVQIIEEHPPMIQNNASGATVATATTFTATSKNSGSNS 173

AF326592 EGDYQLVQHEVLCSMTNTYEVLEFLGRGTFGQVVKCWRGTNEIVAIIKILKNHPSYARQG 240
 AF208291 EGDYQLVQHEVLCSMTNTYEVLEFLGRGTFGQVVKCWRGTNEIVAIIKILKNRPSYARQG 240
 AF207702 EGDYQLVQHEVLCSMTNTYEVSEFLGRGTFGQVVKCWRGTNEIVAIIKILKNHPSYARQG 233
 *****:*****

AF326592 QIEVSILARLSTESADDYNEFRAYECFQHKHNTCLVFEMLEQNLYDFLKQNKFSPLPLKY 300
 AF208291 QIEVSILARLSTESADDYNEFRAYECFQHKHNTCLVFEMLEQNLYDFLKQNKFSPLPLKY 300
 AF207702 QIEVSILARLSTESADDYNEFRAYECFQHKHNTCLVFEMLEQNLYDFLKQNKFSPLPLKY 293

AF326592 IRPVLQVQVATALMKLSLGLIHADLKPENIMLVDPSPRQPYRVKVIDEFSASHVSKAVCST 360
 AF208291 IRPVLQVQVATALMKLSLGLIHADLKPENIMLVDPSPRQPYRVKVIDEFSASHVSKAVCST 360
 AF207702 IRPVLQVQVATALMKLSLGLIHADLKPENIMLVDPSPRQPYRVKVIDEFSASHVSKAVCST 353

AF326592 YLQSRYYRAPEIILGLPFCEAIDMWSLGCVIAELFLGWPLYPGASEYDQIRYISQTQGLP 420
 AF208291 YLQSRYYRAPEIILGLPFCEAIDMWSLGCVIAELFLGWPLYPGASEYDQIRYISQTQGLP 420
 AF207702 YLQSRYYRAPEIILGLPFCEAIDMWSLGCVIAELFLGWPLYPGASEYDQIRYISQTQGLP 413

AF326592 AEYLLSAGTKTTRFFNRDTSFYLWRLKTPDDHEAETGIKSKEARKYIFNCLDDMAQVN 480
 AF208291 AEYLLSAGTKTTRFFNRDTSFYLWRLKTPDDHEAETGIKSKEARKYIFNCLDDMAQVN 480
 AF207702 AEYLLSAGTKTTRFFNRDTSFYLWRLKTPDDHEAETGIKSKEARKYIFNCLDDMAQVN 473

AF326592 MTTDLEGSDMLVEKADRREFIDLLKKMLTIDADKRITPIETLNHPFVTMTHLLDFPHSTH 540
 AF208291 MTTDLEGSDMLVEKADRREFIDLLKKMLTIDADKRITPIETLNHPFVTMTHLLDFPHSTH 540
 AF207702 MTTDLEGSDMLVEKADRREFIDLLKKMLTIDADKRITPIETLNHPFVTMTHLLDFPHSTH 533

AF326592 VKSCFQNMIEICKRRVNMVDTVNQSKTPFITHVAPSTSTNLMTFNNQLTTVHN----- 593
 AF208291 VKSCFQNMIEICKRRVNMVDTVNQSKTPFITHVAPSTSTNLMTFNNQLTTVHNQAPSSST 600
 AF207702 VKSCFQNMIEICKRRVNMVDTVNQSKTPFITHVAPSTSTNLMTFNNQLTTVHNQAPSSST 593

AF326592 -----QPSAASMAAQAQSMPLQTGTQAICARPDFFQQAALIVCPP 633
 AF208291 ATISLANPEVSILNYFSTLYQPSAASMAAQAQSMPLQTGTQAICARPDFFQQAALIVCPP 660
 AF207702 ATISLANPEVSILNYFSTLYQPSAASMAAQAQSMPLQTGTQAICARPDFFQQAALIVCPP 653

AF326592 GFQGLQASPSKHAGYSVRMENAVPIVTQAPGAQPLQIQPGLLAQQAQWPSGTQQILLPPAW 693
 AF208291 GFQGLQASPSKHAGYSVRMENAVPIVTQAPGAQPLQIQPGLLAQQAQWPSGTQQILLPPAW 720
 AF207702 GFQGLQASPSKHAGYSVRMENAVPIVTQAPGAQPLQIQPGLLAQQAQWPSGNQQILLPTW 713
 *****:*****

AF326592 QQLTGVATHTSVQHATVIPETMAGTQQLADWRNTHAHGSHYNPIMQQPALLTGHVTLPA 753
 AF208291 QQLTGVATHTSVQHATVIPETMAGTQQLADWRNTHAHGSHYNPIMQQPALLTGHVTLPA 780
 AF207702 QQLDGVATHTSVQHATVIPETMAGTQQLADWRNTHAHGSHYNPIMQQPALLTGHVTLPA 773
 *** *****

AF326592 QPLNVGVAVHMRQQPTSTTSSRSKQHQSSVRNVSTCEVSSSQAISSPQSRKRVKENTPP 813
 AF208291 QPLNVGVAVHMRQQPTSTTSSRSKQHQSSVRNVSTCEVSSSQAISSPQSRKRVKENTPP 840
 AF207702 QPLNVGVAVHMRQQPTSTTSSRSKQHQ----- 800

AF326592 RCAMVHSSPACSTSVTCGWGDVASSTTRERQRQTIVIPDTPSPTVSVITISSDTDEEEEQ 873
 AF208291 RCAMVHSSPACSTSVTCGWGDVASSTTRERQRQTIVIPDTPSPTVSVITISSDTDEEEEQ 900
 AF207702 -----

```

AF326592      KHAPTSTVSKQRKNVISCVTVHDSFYSDSSNTSPYSVQQRAGHNNANAFDTKGSLENHC 933
AE208291      KHAPTSTVSKQRKNVISCVTVHDSFYSDSSNTSPYSVQQRAGHNNANAFDTKGSLENHC 960
AF207702      -----VSKQRKNVISCVTVHDSFYSDSSNTSPYSVQQRAGHNNANAFDTKGSLENHC 853
                *****

AF326592      TGNPRTIIVPPLKTQASEVLVECDSLVPVNTSHHSSSYKSKSSSNVTSTSGHSSGSSSGA 993
AF208291      TGNPRTIIVPPLKTQASEVLVECDSLVPVNTSHHSSSYKSKSSSNVTSTSGHSSGSSSGA 1020
AF207702      TGNPRTIIVPPLKTQASEVLVECDSLVPGN----- 883
                ***** *

AF326592      ITYRQQRPGPHFQQQQPLNLSQAQQHITTDRTGSHRRQQAYITPTMAQAPYSFPHNSPSH 1053
AE208291      ITYRQQRPGPHFQQQQPLNLSQAQQHITTDRTGSHRRQQAYITPTMAQAPYSFPHNSPSH 1080
AF207702      -----

AF326592      GTVHPHLAAAAAAHLPTQPHLYTYTAPAALGSTGTVAHLVASQGSARHTVQHTAYPASI 1113
AF208291      GTVHPHLAAAAAAHLPTQPHLYTYTAPAALGSTGTVAHLVASQGSARHTVQHTAYPASI 1140
AF207702      -----LG-----EGQGRNLSLESGFPAEL 902
                **          *..* :   . : . : * :

AF326592      VHQVPVSMGPRVLPSPPTIHP SQYPAQFAHQTYISASPASTVYTGYP LSPAKVNQYPYI 1171
AF208291      VHQVPVSMGPRVLPSPPTIHP SQYPAQFAHQTYISASPASTVYTGYP LSPAKVNQYPYI 1198
AF207702      LLEMLLYGS----- 911
                : : : .

```

Figura 8: Comparação da sequência de aa das três isoformas de HIPK2. Em destaque estão os 27aa não encontrados na proteína descrita neste trabalho.

```

1806 TGCAAGCGTC GGGTGAATAT GTATGACACG GTGAACCAGA GCAAAACCCC
      TTTCATCAGC CACGTGGCCC CCAGCACGTC CACCAACCTG ACCATGACCT
      TTAACAACCA GCTGACCACT GTCCACAACC AGGtaagtgg gaaacagaag
      ..... atcccatca ctccttttagG CTCCCTCCTC TACCAGTGCC
      ACTATTTCTT TAGCCAATCC CGAAGTCTCC ATACTAAACT ACCCATCTAC
      ACTCTACCAG CCCTCAGCGG CATCCATGGC TGCAGTGGCC CAGCGGAGCA
      TGCCCCTGCA GACAGGAACA GCCCAGATTG GTGCCCGGCC TGACCCGTC
      CAGCAAGCTC TCATCGTGTG TCCCCCGGCG TTCCAAGGta agattgagca

```

Figura 9: Sequência genômica da região de HIPK2 que se encontra os dois sítios acceptors de splicing que geram as duas isoformas de HID. Em azul estão representados os éxons e em rosa estão indicados os sítios donors. Em amarelo estão destacados os dois possíveis sítios acceptors de splicing e em verde está a sequência gerada pelo primeiro sítio acceptor, que codifica 27aa presentes na isoforma de número de acesso AF208291. A sequência dos íntrons está parcialmente representada em letras pretas e o número indica o início da região de nossa isoforma que está destacada.

4.1d. Avaliação da variação de alaninas no domínio YH em diferentes espécies

Comparando-se as sequências de aminoácidos de HIPK2 com aquelas de camundongo e hamster, encontramos no domínio YH uma alta identidade, onde apenas o número de alaninas variava entre as duas espécies. No intuito de avaliar esta variação em outras espécies de mamíferos, amplificamos esta região usando DNA de diferentes espécies. Os produtos obtidos foram sequenciados e avaliados com o auxílio do programa *GeneRunner*

e do programa de alinhamento BLASTX. As amostras tamanduá bandeira, tamanduá mirim, tatu, furão, mão pelada, cachorro do mato e gato do mato foram cedidas por José Eduardo Garcia e Profa Eucleia Primo Betioli Contel do Departamento de Genética da Faculdade de Medicina de Ribeirão Preto, SP. As amostras de preguiça, gambá, mico, cavalo e cão foram cedidas pelo Prof. Dr. Fabrício Rodrigues dos Santos, do Departamento de Biologia Geral do Instituto de Ciências Biológicas da UFMG. A amostra de galinha foi cedida pelo Prof. Dr. Luiz L. Coutinho, ESALQ/USP, Piracicaba, SP. As amostras de porco e touro foram obtidas em açougue. A tabela 6 mostra os animais dos quais conseguimos seqüenciar esta região. Não foi possível obter todas as seqüências das espécies citadas acima devido a baixa qualidade do DNA e/ou a escassez do material para as padronizações com as diferentes duplas de iniciadores (descritos no tópico 3.3 da seção Materiais e Métodos). Esta padronização consumiu um grande tempo, sendo necessário a amplificação com taq Platinum Hi-Fi (GibcoBRL) que possui *hot start*.

Nome popular	Nome científico	Ordem	Obtido através:	Número de alaninas repetidas no domínio YH
Tamanduá-bandeira	<i>Myrmecophaga tridactyla</i>	Xenarthra	José Eduardo Garcia e Profa Eucleia Primo Betioli Contel, Depto de Genética da Faculdade de Medicina de Ribeirão Preto, SP	6
Tamanduá-mirim	<i>Tamandua tetradactyla</i>	Xenarthra	José Eduardo Garcia e Profa Eucleia Primo Betioli Contel, Depto de Genética da Faculdade de Medicina de Ribeirão Preto, SP	6
Preguiça comum	<i>Bradypus variegatus</i>	Xenarthra	Prof. Dr. Fabrício Rodrigues dos Santos, Depto de Biologia Geral do Inst. De Ciências Biológicas da UFMG	3
Gambá	<i>Didelphis marsupialis</i>	Marsupialia	Prof. Dr. Fabrício Rodrigues dos Santos, Depto de Biologia Geral do Inst. De Ciências Biológicas da UFMG	4
Gato-do-mato	<i>Felis tigrina</i>	Carnívora	José Eduardo Garcia e Profa Eucleia Primo Betioli Contel, Depto de Genética da Faculdade de Medicina de Ribeirão Preto, SP	7
Cachorro-do-mato	<i>Cerdocyon thous</i>	Carnívora	José Eduardo Garcia e Profa Eucleia Primo Betioli Contel, Depto de Genética da Faculdade de Medicina de Ribeirão Preto, SP	7
Mico-leão	<i>Leontopithecus sp</i>	Primatas	Prof. Dr. Fabrício Rodrigues dos Santos, Depto de Biologia Geral do Inst. De Ciências Biológicas da UFMG	6
Porco	<i>Sus scrofa domestica</i>	Suina	Açougue	3
Cavalo	<i>Equus caballus</i>	Perissodactyla	Prof. Dr. Fabrício Rodrigues dos Santos, Depto de Biologia Geral do Inst. De Ciências Biológicas da UFMG	6
Cachorro	<i>Canis familiaris</i>	Carnívora	Prof. Dr. Fabrício Rodrigues dos Santos, Depto de Biologia Geral do Inst. De Ciências Biológicas da UFMG	7
Galinha	<i>Gallus gallus domesticus</i>	Galliformes	Prof. Dr. Luiz L. Coutinho, ESALQ/USP, Piracicaba, SP	6

Tabela 6: Animais utilizados no estudo da região polimórfica do domínio YH.

A figura 10 ilustra alguns produtos de PCR de diferentes espécies. Foi necessário também padronizar as reações de sequenciamento. Nesta fase contamos com a importante colaboração da Dra. Maria Cristina Ramos Costa, da Fundação Hemocentro de Ribeirão Preto que realizou parte do sequenciamento.

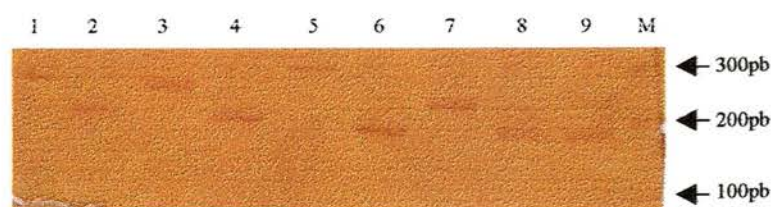
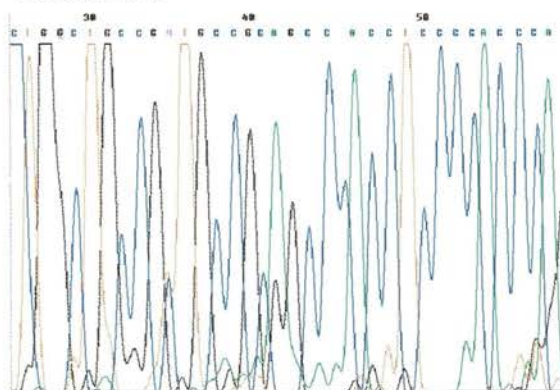


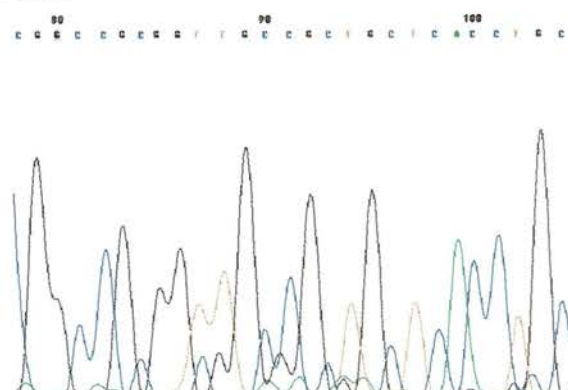
Figura 10: PCR da região codificadora das repetições de alaninas de HIPK2 em diferentes animais . A amplificação foi realizada com diferentes combinações de iniciadores e DNAs derivados de tamanduá mirim (1), tamanduá bandeira (2), gato do mato (3 e 4), furão (5), porco (6 e 7) e gambá (8 e 9). M representa o marcador de peso molecular 100pb (Gibco BRL) e as setas indicam o tamanho das bandas de M.

A figura 11 mostra as seqüências obtidas pelo seqüenciamento das bandas derivadas de DNA de tamanduá bandeira, tamanduá mirim, preguiça, gambá, cachorro do mato, gato do mato, porco, touro e galinha. A figura 12 compara as seqüências de aminoácidos das diferentes espécies em conjunto e as proteínas do camundongo (Hipk2), do hamster (PKM) e humana.

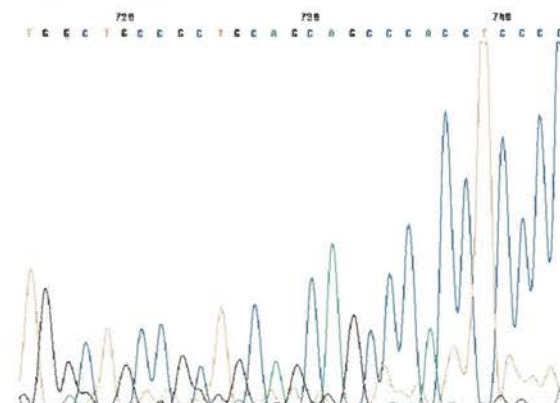
Tamandua Mirim



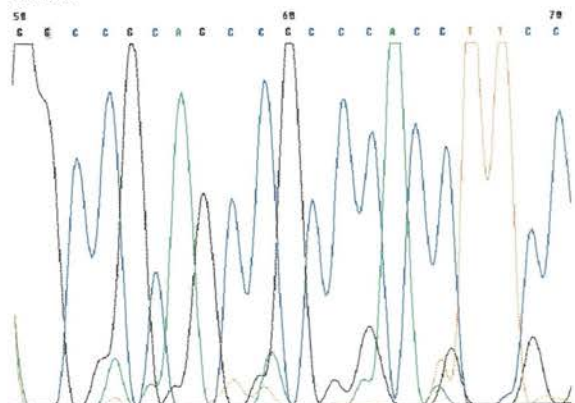
Galinha



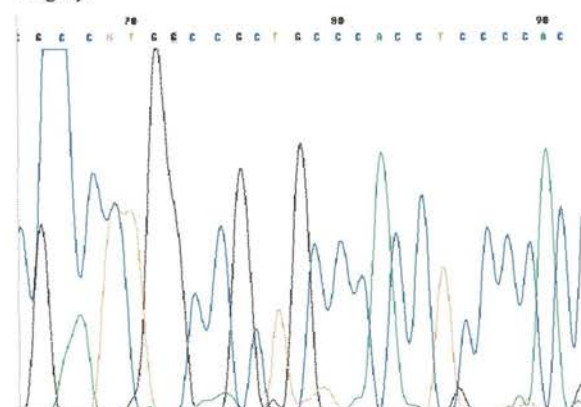
Tamandua Bandeira



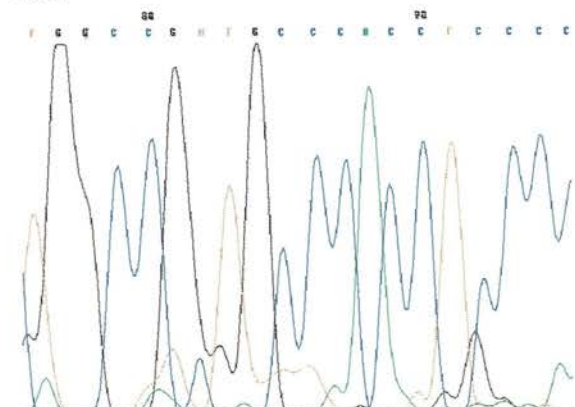
Gambá



Preguiça



Porco



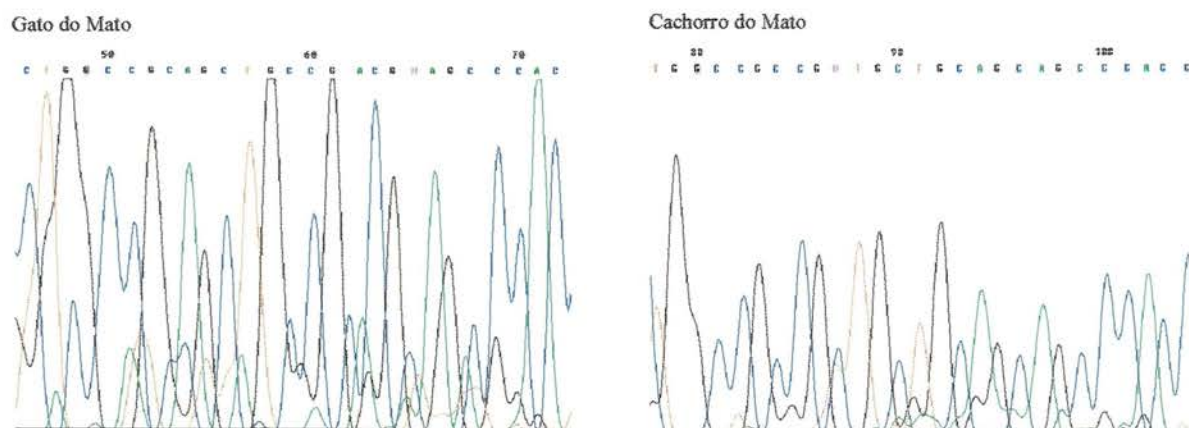


Figura 11: Sequências dos diferentes animais

```

>HIPK2 1083 VHPHLAAAAAAHLPTQPHLYTYTAPAALGSTGTVAHLVASQGSARHTVQH
>Hipk2 1049 VHPHL AAAAHLPTQPHLYTYTAPTALGSTGTVAHLVASQGSARHTVQH
>Mx-PKM 1051 VHPHL AAAAHLPTQPHLYTYTAPTALGSTGTVAHLVASQGSARHTVQH
>Gam1 HL AAAAHLPAQPHLYTYTAPAALGSSGTVAHLVASQGSARHAVQH
>Porc VHPHL AAAHLPTQPHLYTYTAPAALGSTGTVAHLVASQGSARHTVQH
>Preg HL AAAHLPTQPHLYTYTAPAALGSTGTVAHLVASQGXAR
>Gal est AAAAHLPTQPHLYTYTTPAALGSTGTVAHLVASQGSARHAVQH
>Gal HP AAVAAHLPTQPHLYTYTTPAALGSTGTVAHL
>Tmir YL AAAAAHLPTQPHLYTYTAPAALGSTGTVAHLVASQGSARHTVQH
>Tband HL AAAAAHLPTQPHLYTYTAPAALGSTGTVAHLVASQGS
>Cavalo VHPHL AAAAAHLPAQPHLYTYTAPAALGSTGTVAHLVASQGS
>Mico VHPHL AAAAAHLPTQPHLYTYTAPAALGSTGTVAHLVASQGSAXHTVQH
>Mico VHPHL AAAAAHLPTQPHLYTYTAPAALGSTGTVAHLVASQGSARHTVQH
>Cão LAAAAAHLPAQPHLYTYTAPAALGSTGTVAHLVASQGSARHTVQH
>CacM HLAAAAAHLPTQPHLYTYTAPAALGSTGTVAHLGASQGSARHTVQH
>GatM HLAAAAAHLPTQPHLYTYTAPAALGSTGTVAHLVASQGSARHTVQH

>HIPK2 TAYPASIVHQVPVSMGPRVLPSPITIHPSQYPAQFAHOTYISASPASTVYTG
>Hipk2 TAYPASIVHQVPVSMGPRVLPSPITIHPSQYPAQFAHOTYISASPASTVYTG
>Mx-PKM TAYPASIVHQVPVSMGPRVLPSPITIHPSQYPAQFAHOTYISASPASTVYTG
>Gam1 TAYPASI
>Porc1 TAYPASIVHQVPVSMGPRVLPSPITIH
>Gal est TTYPASIVHQVPVSMGPRVLPSPITIHPSQYPAQFAHOTYISASPASTVYTG
>Tmir1 TAYPASIVHQVPVSMGPRVLPSPITIH
>Mico TAYPASIVHQVPVSMAP
>Mico TAYPASIVHQVPVSMGPRVLPXPT
>Cão TAYPASIVHQVPVSMGPRVLPSPITIHPSQY
>CacM TAYPASIVHQVPVSMGPRVLPSPITIH
>GatM TAYPASIVHQVPVSM

>HIPK2 YPLSPAKVNQYPYI 1198
>Hipk2 YPLSPAKVNQYPYI 1161
>Mx-PKM YPLSPAKVNQYPYI 1162
>Gal est YPLSPTKVNQYPYI

```

Figura 12: Sequência de aminoácidos de HIPK2 comparada com a sequência de camundongo (Hipk2), hamster (Mx-PKM) e os animais sequenciados no projeto. As sequências das diferentes espécies foram traduzidas pelo programa BLASTX e GeneRunner para serem comparadas com as sequências de humano e de roedores já existentes. Os animais sequenciados foram gambá (Gam), galinha (Gal e também a sequência depositada no banco de dados, Gal est), porco (Porc), preguiça (Preg), tamanduá mirim (Tmir), tamanduá bandeira (Tband), cavalo, mico, cão, cachorro do mato (CacM) e gato do mato (GatM). As letras sublinhadas em cinza são idênticas à sequência da proteína humana.

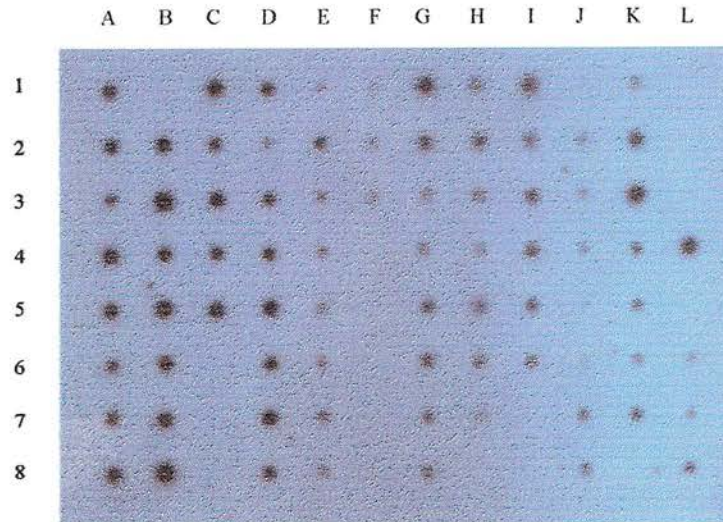
4.2. Análise da expressão deste gene

4.2a. Northern Blot

Para avaliar e comparar a expressão do gene *HIPK2* em diferentes tecidos normais e tumorais preparamos uma membrana contendo 15µg de RNA total extraído de tecido tumoral e normal (adjacente ao tumor) de cólon (2 amostras), mama (2 amostras) e língua (2 amostras). Para comprovar a qualidade do RNA da membrana, esta foi hibridizada, segundo o protocolo descrito por Church e Gilbert, 1984, com uma sonda derivada de GAPDH, cedida pelo laboratório do Dr. Luis Fernando Reis também do Instituto Ludwig de Pesquisa contra o Câncer. A sonda de *HIPK2* utilizada foi obtida pela digestão do clone AA4553887 do IMAGE, liberando um inserto de aproximadamente 1500pb que foi purificado utilizando o *kit QIAquick™ Gel Extraction* (Quiagen) e quantificado em gel de agarose 1% utilizando como padrão de bandas o *Low Mass Ladder* (GibcoBRL). A sonda foi então marcada e a hibridização foi feita também segundo o protocolo descrito por Church e Gilbert, 1984. O resultado obtido foi uma possível marcação inespecífica no RNA, com as regiões dos RNAs ribossomais 18S e 28S com menor marcação. O experimento foi repetido utilizando uma nova membrana de RNA total com diferentes amostras das mesmas regiões utilizadas no primeiro experimento encontrando-se o mesmo tipo de hibridização inespecífica.

Realizamos então esta análise de *Northern Blot* utilizando uma membrana comercial (Clontech), contendo 12 amostras de mRNA de tecidos humanos, utilizando a mesma sonda previamente usada nas membranas de RNA total. A figura 13 mostra o resultado após exposição no aparelho *Storm 840*, Molecular Dynamics, por uma semana. Três formas de *splicing* foram encontradas, uma com mais de 10 kb, uma com aproximadamente 7 kb e outra com aproximadamente 4 kb, que corresponde à forma que obtivemos a partir de mRNA de cérebro adulto. A visualização das bandas só foi possível nos tecidos derivados de cérebro, coração, músculo esquelético, rim e fígado, onde dados obtidos por *Dot Blot* indicam maior expressão (a ser comentado no próximo tópico).

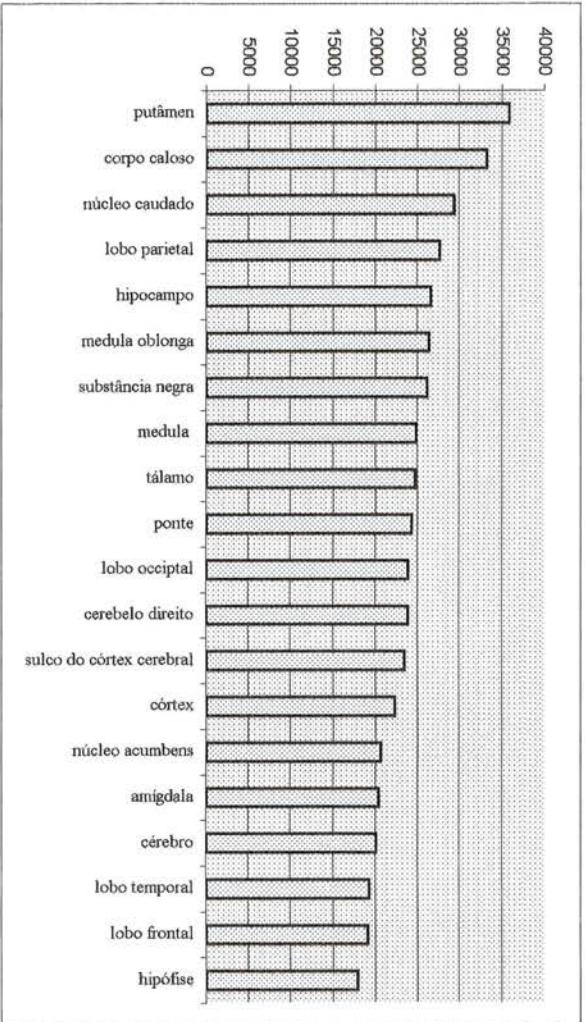
a



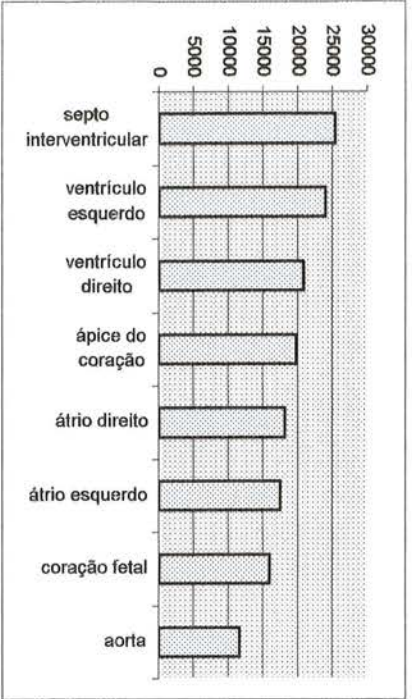
b

	A	B	C	D	E	F	G	H	I	J	K	L
1	Cérebro total		Substância negra	Coração	Esôfago	Cólon transverso	Rim	Pulmão	Fígado	Leucemia HL-60	Cérebro fetal	RNA total levedura
2	Cortex	Cerebelo direito	Núcleo acumbens	Aorta	Estômago	Cólon descendente	Músculo esquelético	Placenta	Pâncreas	Hela S3	Coração fetal	tRNA levedura
3	Lobo frontal	Corpo caloso	Tálamo	Átrio esquerdo	Duodeno	Reto	Baço	Bexiga	Glândula adrenal	Leucemia K-562	Rim fetal	RNA <i>E. coli</i>
4	Lobo parietal	Amígdala	Hipófise	Átrio direito	Jejuno		Timo	Útero	Tireóide	Leucemia MOLT-4	Fígado fetal	DNA <i>E. coli</i>
5	Lobo occipital	Núcleo caudado	Medula	Ventrículo esquerdo	Íleo		Linfócitos do sangue periférico	Próstata	Glândula salivar	Linfoma de Burkitt's a Raji	Baço fetal	Poli r (a)
6	Lobo temporal	Hipo-campo		Ventrículo direito	Íleo/ceco		Linfonodo	Testículo	Glândula mamária	Linfoma de Burkitt's a Daudi	Timo fetal	DNA humano C _{ot} -1
7	Sulco	Medula oblonga		Septo inter-ventricular	Apêndice cecal		Medula óssea	Ovário		Adeno-carcinoma colorctal SW480	Pulmão fetal	100ng DNA humano
8	Ponte	Putâmen		Ápice do coração	Cólon ascendente		Traquéia			Carcinoma pulmonar A549		500ng DNA humano

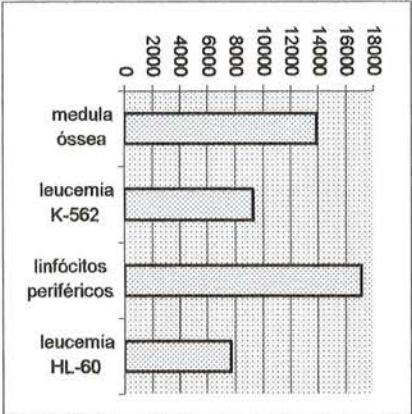
Figuras 14 a e b: Análise da expressão de HIPK2 por Dot Blot. A membrana Human Multiple Tissue Expression (MTETTM) Array (Clontech) com 2µg de mRNA de 76 amostras de tecidos humanos (b), hibridizada com a sonda da região 3' de HIPK2 e foi então exposta a um filme de Raio-X a -70°C por 6 dias(a).



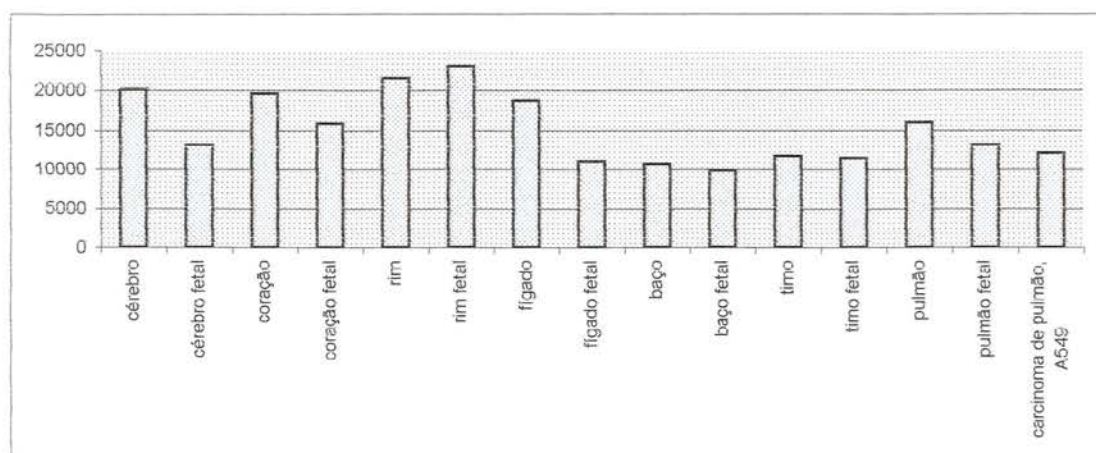
1



2



3



Gráficos 1, 2, 3 e 4: Comparação dos diferentes níveis de expressão obtidos pelo Dot Blot. O gráfico 1 compara a expressão entre diferentes tecidos cerebrais, o gráfico 2 entre diferentes regiões do coração e também inclui a aorta, o gráfico 3 entre diferentes linhagens de leucemias e seu tecido equivalente normal e o gráfico 4 compara diferentes tecidos, adulto e fetal, e quando existente, a linhagem tumoral correspondente. A intensidade do sinal de hibridização é apresentada em pixels, no eixo das ordenadas.

4.2c. Análises de SAGE (Serial Analysis of Gene Expression)

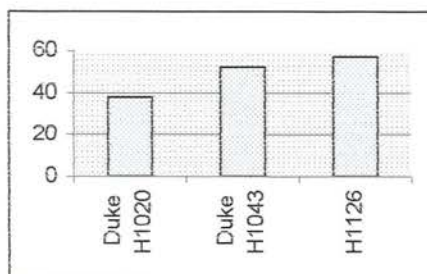
A tabela 7 mostra a análise de SAGE (Serial Analysis of Gene Expression, <http://www.ncbi.nlm.nih.gov/sage>) deste gene atualizada em 19 de novembro de 2001. O transcrito de *HIPK2* foi encontrado em 23 das 100 bibliotecas disponíveis até o momento. A maioria destas bibliotecas foram derivadas de cérebro (10/23). Em algumas amostras comparamos o número de SAGE tags em dois diferentes graus de diferenciação de um mesmo tumor ou em situação de hipóxia de uma mesma linhagem celular. O gráfico 5 mostra a comparação de duas amostras obtidas por SAGE de astrocitomas, um bem diferenciado e o outro anaplásico. O gráfico 6 mostra a mesma linhagem celular de glioma submetida ou não a hipóxia.

Em paralelo as análises de SAGE procuramos o número de ESTs depositadas no banco de dados de ESTs que representam o transcrito de *HIPK2*. Para isto utilizamos o banco de seqüências do UNIGENE (<http://www.ncbi.nlm.nih.gov>) que nos mostrou a existência de 92 ESTs (número de acesso de *HIPK2* no UNIGENE é Hs.236131).

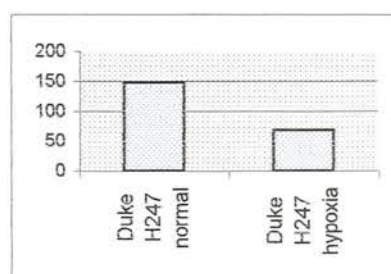
		milhão	Tags total
SAGE OVP-5	Linhagem Celular (LC) tumoral De ovário	277	3/10.802
SAGE Duke 247 normal	LC de glioblastoma multiforme	148	9/60.543
SAGE Duke precrisis fibroblast	LC de células da pele transformadas com antígeno T que não atingiram o estado de crise	113	1/8.797
SAGE Duke H392	LC de glioblastoma multiforme	104	6/57.529
SAGE BB542 withematter	Amostra de cérebro normal com mais de 95% de substância branca	73	7/94.806
SAGE Duke H988	Oligodendroglioma Bem diferenciado	71	2/28.015
SAGE Duke H247 Hypoxia	LC de glioblastoma multiforme crescido em condições de hipóxia	69	5/71.937
SAGE H1126	Astrocitoma fibrilar bem diferenciado	57	1/17.501
SAGE H1043	Astrocitoma Pilocítico	52	4/76.673
SAGE SW837	LC tumoral de cólon com p53 e APC mutantes	49	3/60.986
SAGE LN-1	Tumor avançado de próstata	43	1/22.835
SAGE 293-IND	LC 293 derivada de rim embrionário Expressando beta-catenina	40	1/24.481
SAGE Duke H1020	Astrocitoma Anaplásico	38	2/52.371
SAGE A+	LC tumoral de próstata induzida com androgênio sintético	32	1/30.551
SAGE HS766T	LC tumoral De pâncreas	31	1/31.506
SAGE ES2-1	LC de carcinoma de células claras de ovário	31	1/31.502
SAGE Duke Kidney	Rim adulto normal	24	1/41.413
SAGE CPDR LNCaP-T	LC de carcinoma de próstata	22	1/44.122
SAGE mammary epithelium	Epitélio luminal da mama	20	1/49.167
SAGE normal cerebellum	Cerebelo Normal	19	1/51.135
SAGE RKO	LC tumoral de cólon com p53 mutante	19	1/52.064
SAGE Duke H54 EGFRvIII	LC de glioblastoma multiforme transfectada com o receptor mutado do fator de crescimento epidermal	17	1/57.164
SAGE SciencePark MCF7 estradiol 10h	LC de carcinoma mamário Tratado com estradiol	16	1/60.435

Tabela 7: Resultados da análise de SAGE

5



6



Gráficos 5 e 6: Comparação dos dados obtidos por SAGE. O gráfico 5 compara a expressão virtual de HIPK2 em 3 astrocitomas, sendo Duke H1020 anaplásico e Duke H1043 e H1126, bem diferenciados. O gráfico 6 compara a expressão virtual da mesma linhagem celular de glioblastoma, submetida ou não a hipóxia.

4.2d. Hibridização *in situ* de HIPK2 em camundongo fetal

Com o objetivo de avaliar a expressão específica deste gene em alguns tecidos foram realizadas hibridizações *in situ* em camundongos, em colaboração com o Prof. Dr. Fábio Daumas Nunes, do laboratório de Biologia Molecular do Instituto de Odontologia da Universidade de São Paulo. Diante da dificuldade em se obter resultados dirigidos aos órgãos de interesse onde o gene mostrou maior expressão (rim, coração e cérebro de animais adultos), a hibridização foi realizada em cortes seriados de fetos de 13 dias pós-coito (dpc), no estágio 21, segundo a classificação de Theiler (1989). O resultado da hibridização é mostrado na figura 15.

Os dados obtidos mostram, em camundongos do estágio 21, um sinal de hibridização intenso e bem delimitado nos tecidos epiteliais que recobrem a faringe, traquéia, esôfago e intestino delgado (fig15 B, C e E). Nos pulmões em desenvolvimento, os brônquios principais, primários e segmentares também apresentaram sinal de hibridização (figura 15 C). Um sinal bem concentrado foi observado na metade dorsal do lúmen estomacal (figura 15 D). Um sinal menos intenso, porém distinto foi visto nos ductos mesonéfricos. Ao longo de toda extensão rostro-caudal do camundongo fetal foi observada uma marcação intensa no tubo neural, principalmente em suas células ventrais (figura 15 A). A notocorda também foi marcada. Não se observou nenhuma marcação no coração.

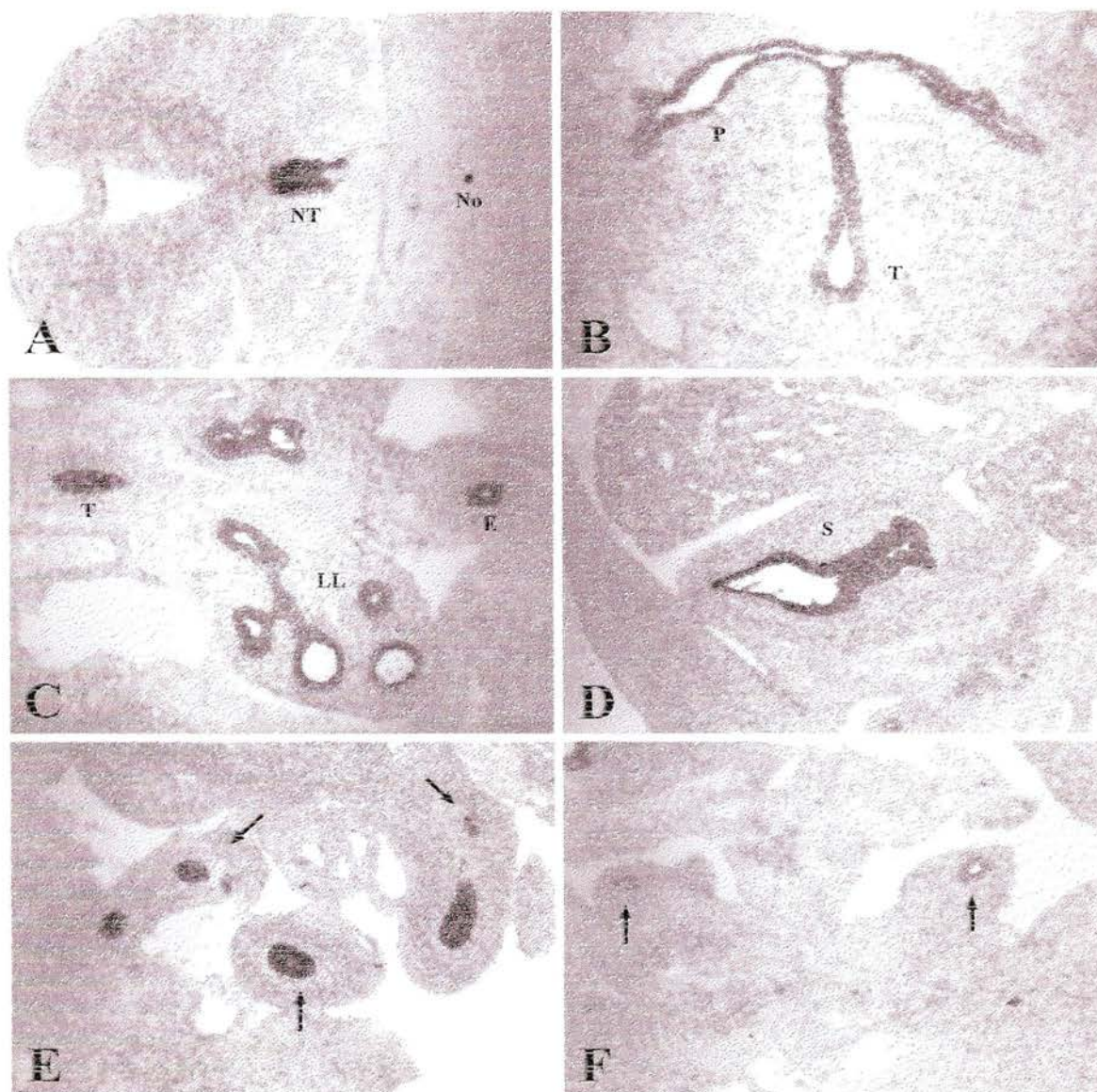


Figura 15: Hibridização *in situ* de *HIPK2*. A expressão *HIPK2* foi detectada (A) no tubo neural (NT) e na notocorda (No). A hibridização do NT foi observada em toda a extensão rostro-caudal do embrião no estágio 21 do desenvolvimento. (B) A borda epitelial da faringe (P) e traquéia (T) também mostraram hibridização com *HIPK2*. (C) A traquéia foi hibridizada de sua extensão ventral até a extensão mais caudal. Os pulmões também mostraram hibridização (LL), junto com o esôfago (E). (D) A expressão de *HIPK2* foi bastante concentrada na metade dorsal do epitélio estomacal (S). (E) Encontrou-se este transcrito no intestino delgado (E) e nos ductos mesonéfricos (F). As fotos foram tiradas a partir das lâminas ampliadas 100X do seu tamanho original.

4.2e. Análise das formas de *splicing* do domínio de interação com as homeoproteínas (HID)

I. Análises das isoformas de HID no desenvolvimento do camundongo

Diante dos achados de *HIPK2* na hibridização *in situ*, e também diante da existência de pelo menos duas formas de *splicing* em uma importante região da molécula, resolvemos avaliar a expressão destas em alguns estágios do desenvolvimento e após o nascimento, utilizando RT-PCR. Contamos mais uma vez com a colaboração do Prof. Dr. Fábio Daumas Nunes, que auxiliou na obtenção de material biológico, responsabilizando-se pela dissecação e coleta dos órgãos de interesse (coração, cérebro e rim). Utilizamos camundongos em dois estágios fetais (15dpc e 17 dpc), três dias após o nascimento, 5 dias após o nascimento e camundongos adultos. Utilizamos um par de iniciadores (D e T) que permite amplificar uma porção do HID, na qual verificamos diferentes formas de *splicing*.

O produto desta PCR com os iniciadores D e T gerou três bandas de tamanhos moleculares distintos, duas já de tamanho esperado. A banda 1, que corresponde à isoforma caracterizada neste trabalho, com o tamanho esperado de 469pb, a banda 2, que representa a isoforma descrita em março de 2001, com 550pb e uma terceira banda de uma possível isoforma ainda não caracterizada com aproximadamente 1.050pb. Esta terceira isoforma é de visualização mais difícil em géis de agarose, sistema que usamos preferencialmente nestes experimentos por ser mais adequado para análises de quantificação. Em paralelo a este experimento avaliamos esta mesma região em alguns tumores humanos (este tópico será abordado posteriormente), onde foram encontradas as mesmas três isoformas descritas neste experimento.

Para melhor analisarmos a expressão destas diferentes isoformas de *HIPK2* nas distintas fases do camundongo, foi feita uma PCR semi-quantitativa com 21, 28, 35 e 40

ciclos. A figura 16 ilustra nossos resultados em rim. A terceira isoforma, a maior delas, não foi avaliada por ser mais aparente em géis de poliacrilamida corados por prata.

Os resultados obtidos indicam que no rim a banda 2 aparenta ser mais intensa nos estágios fetais, diminuindo esta intensidade nos estágios pós-parto, com cinco dias após o nascimento esta intensidade se iguala entre as bandas 1 e 2 e no camundongo adulto a primeira banda prevalece sobre a segunda.

No cérebro (figura 17), no estágio de 15dpc, as duas isoformas representadas pela banda 1 e banda 2 parecem ter a mesma intensidade. No estágio 17dpc há um aumento da expressão do transcrito representado pela banda 2. Nos dias 3 e 5 após o nascimento há prevalência da banda 1 e no adulto a banda 2 quase desaparece como ocorre no rim.

Nos estágios fetais do coração (figura 17) a banda 2 é a mais intensa apenas nos três dias após o nascimento e no quinto dia a intensidade das duas bandas se iguala e se mantém no coração adulto.

Os dados da expressão em diferentes estágios do desenvolvimento do cérebro e coração estão ilustrados na figura 17, que contém os diferentes estágios dos órgãos alvos e a amplificação conjunta das bandas específicas com iniciadores específicos para *GAPDH*, um gene constitutivo, como controle da quantidade do cDNA de cada tecido.

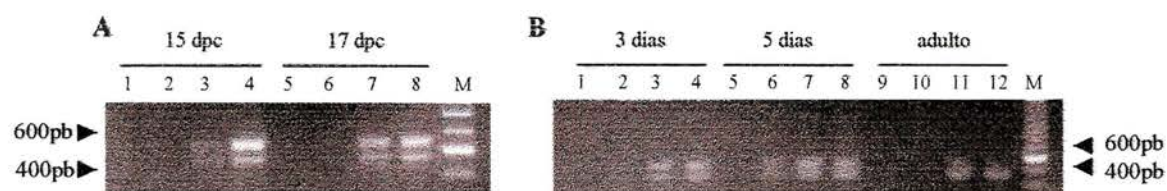


Figura 16: PCR semi-quantitativa de *HIPK2* em diferentes estágios do desenvolvimento renal de camundongos. As figuras A e B mostram o resultado da eletroforese em gel de agarose 1,5% com o produto da amplificação de *HIPK2* em rim de camundongos com 21 ciclos (1A, 5A, 1B, 5B, 9B), 28 ciclos (2A, 6A, 2B, 6B, 10B), 35 ciclos (3A, 7A, 3B, 7B, 11B) e 40 ciclos de (4A, 8A, 4B, 8B, 12B). As canaletas 1A a 4A representam os rins de fetos com 15 dias pós coito (dpc), 5A a 8A rins de fetos com 17dpc, 1B a 4B, rins de camundongos com 3 dias pós-parto, 5B a 8B, rins de camundongos com 5 dias pós-parto e 9B a 12B rins de camundongos adultos. M representa o marcador de peso molecular 100pb (GibcoBRL).

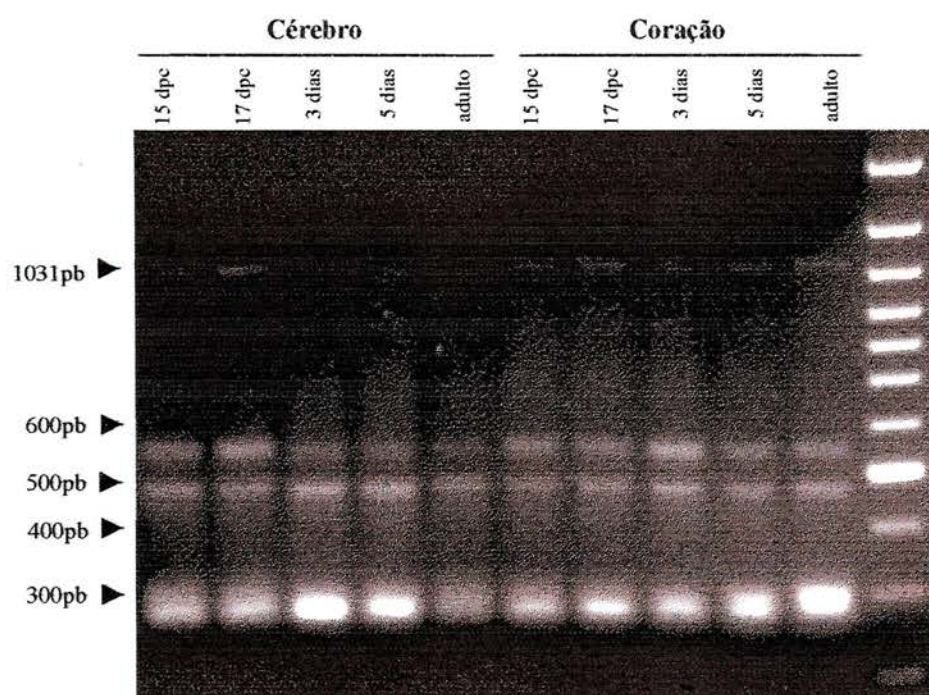


Figura 17: PCR de *HIPK2* em diferentes estágios do desenvolvimento de cérebro e coração de camundongos. As canaletas 1 a 5 representam a amplificação por 40 ciclos de cDNAs derivados de RNAs de cérebros de camundongos em diferentes estágios de desenvolvimento. Os períodos avaliados correspondem a 15 dias pós-coito (15 dpc) canaleta 1; 17dpc canaleta 2); três dias após o nascimento, canaleta 3; 5 dias após o nascimento, canaleta 4 e animal adulto na canaleta 5. As canaletas 6 a 10 seguem estes mesmos estágios de desenvolvimento, mas do coração destes camundongos. A banda de aproximadamente 300pb é a do gene constitutivo *GAPDH*. M representa o marcador de peso molecular 100pb (Fermentas) e as setas indicam o tamanho das bandas de M.

II. Análise das isoformas de HID em tecidos e linhagens celulares humanos

Paralelamente aos estudos de expressão de *HIPK2* em diferentes estágios do desenvolvimento, avaliamos a sua expressão e a presença de formas alternativas de *splicing* utilizando o mesmo par de iniciadores usado no tópico anterior (D e T). A figura 18 mostra os resultados obtidos em alguns tecidos, normais ou tumorais e em uma linhagem de carcinoma de cólon.

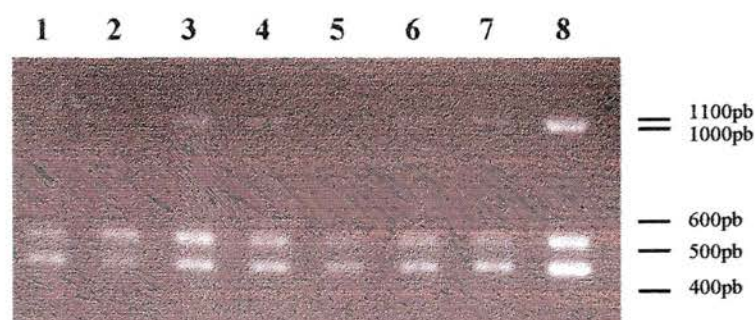


Figura 18: Expressão das diferentes isoformas de *HIPK2* em tecidos humanos. Com nanogramas de mRNA de tecidos derivados de língua normal (1) e tumoral (2), cólon normal (3) e tumoral (4), células amnióticas (5), tumor de ovário (6), tumor de testículo (7) e linhagem celular de carcinoma de cólon (8) foram usados para RT-PCR do HID. O resultado foi visualizado em gel de agarose 1,5%. M representa o marcador de peso molecular 100pb (GibcoBRL) e os números em pares de base (pb) indicam o tamanho das bandas de M.

O resultado sugere um padrão de expressão de *HIPK2* diferente para os dois conjuntos, normal e tumoral, de língua e cólon. Na primeira dupla parece ocorrer uma inversão de intensidade das bandas, onde a isoforma representada pelo fragmento de 550pb é mais expressa em tumor. Já em cólon o padrão de intensidade das duas bandas parece se manter inalterado, o que também é visto na linhagem celular tumoral de cólon. Para uma confirmação destes resultados seria necessário um número maior de amostras destes tecidos. A principal informação desta figura é o aparecimento da terceira banda em todas

as amostras, confirmando uma terceira forma de *splicing* desta região. Esta nova isoforma pode representar uma das bandas observada no experimento de *Northern Blot*.

III. Análise das isoformas de HIPK2 em linhagens celulares de gliomas de humanos e ratos

No intuito de avaliar a expressão das diferentes isoformas de *HIPK2* em tumores, para caracterizar uma possível relação destas com a progressão tumoral, fizemos uma colaboração com o grupo da Profa. Dra. Mari Sogayar, do departamento de Bioquímica do Instituto de Química da Universidade de São Paulo. O laboratório da Profa. Sogayar trabalha com algumas linhagens de gliomas humanos e de camundongos e o modelo de diferenciação celular com ácido retinóico e glicocorticóides. Primeiramente trabalhamos com amostras de linhagens celulares de gliomas humanos tratadas e não tratadas com ácido retinóico por 48h (figura 19), amostra cedida pela Dra. Regina Maki Sasahara.

O produto da RT-PCR do cDNA derivado da linhagem A172 após o tratamento com ácido retinóico, tem a intensidade de sua banda 2 reduzida. Este padrão de resposta não ocorre na linhagem T98G. Nestas amostras, encontramos novamente a terceira banda com aproximadamente 1.050pb.

Para uma melhor caracterização de *HIPK2* e a expressão de suas diferentes isoformas na progressão tumoral a aluna Diana Nunes, do laboratório da Profa. Dra. Mari Sogayar, nos forneceu amostras de RNA de duas variantes da linhagem de glioma de ratos C6, a variante ST1 e a P7. A variante ST1, quando tratada com hidrocortisona reverte seu fenótipo, ou seja se diferencia, já a variante P7 não responde ao tratamento.

Após a transcrição reversa dos RNAs extraídos de ambas variantes com diferentes tempos de tratamento e a amplificação com os iniciadores específicos para a região que

apresenta as três isoformas, foi possível avaliar o padrão de expressão destas. A terceira isoforma novamente não foi considerada, como já citado no tópico 2 (figura 20).

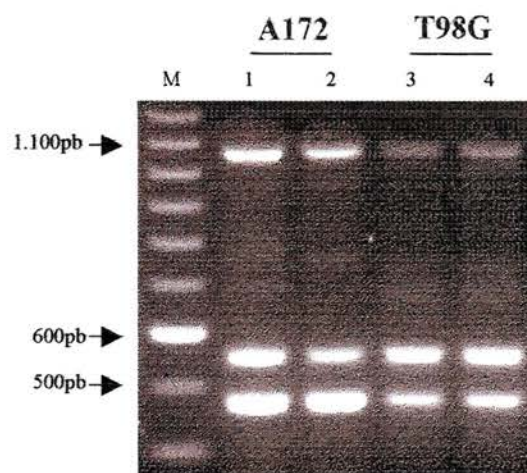


Figura 19: Expressão das diferentes isoformas de HIPK2 em linhagens celulares de gliomas humanos. cDNA de duas linhagens celulares de glioma, a A172 (canaletas 1 e 2) e a T98G (canaletas 3 e 4) foram usados para amplificação do HID. Estas linhagens foram tratadas (2 e 4) por 48h com ácido retinóico. Canaletas 1 e 3 mostram as linhagens sem tratamento. O resultado foi visualizado em gel de agarose 1,5%. M representa o marcador de peso molecular 100pb (GibcoBRL) e as setas indicam o tamanho das bandas de M.

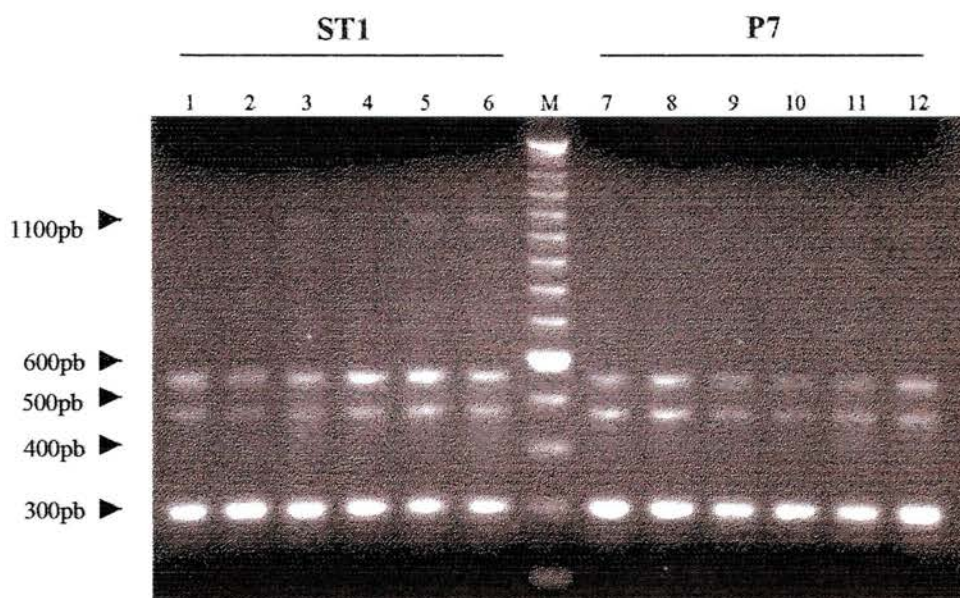


Figura 20: Expressão das diferentes isoformas de HIPK2 nas duas variantes da linhagem C6 de gliomas de ratos tratadas com hidrocortisona em diferentes tempos. A amplificação foi realizada com cDNAs derivados de RNAs das diferentes variantes ST1 (1-6) e P7 (7-12) tratadas com hidrocortisona por 30 minutos (2 e 8), 2 horas (3 e 9), 8h (4 e 10), 24h (5 e 11) e 48h (6 e 12). As canaletas 1 e 7 representam o controle do tratamento, ou seja, as variantes sem tratamento. A banda de aproximadamente 300pb é a do gene constitutivo *GAPDH*. M representa o marcador de peso molecular 100pb (Gibco BRL) e as setas indicam o tamanho das bandas de M.

A variante ST1 parece ter uma diminuição da expressão das duas isoformas (banda 1 e 2) com 30 minutos de tratamento com hidrocortisona, mas no controle a segunda isoforma (banda 2) tem uma expressão maior do que a primeira. A partir de 2 horas de tratamento as duas bandas se tornam mais fortes, mas a banda 2 tem sempre uma maior intensidade, sendo mais forte com 24 horas de tratamento e a partir de 48 horas esta intensidade parece voltar para seu nível anterior ao tratamento. Na variante P7 as duas bandas tem a mesma intensidade. A partir de 8 horas de tratamento as duas bandas tendem a diminuir sua intensidade durante todo o tratamento.

4.3. Análises da possível terceira isoforma de HIPK2

O produto da RT-PCR da região amplificada pelos iniciadores D e T, quando visualizados em gel de agarose, originava 3 bandas de tamanhos distintos, duas destas representavam a isoforma descrita neste trabalho e a outra isoforma, já publicada (Hoffman et al., 2000), com 81nt a mais que a anterior. A terceira banda, de aproximadamente 1.050pb, representa uma possível nova isoforma no início do HID. Análises genômicas nesta região apontam a presença de três íntrons, de aproximadamente 5,9Kb, 1,0Kb e 9,0Kb respectivamente (figura 21). O clone genômico CTD-2022G18 (número de acesso no *GenBank*: AC073184) possui esta região e foi usado nas análises.

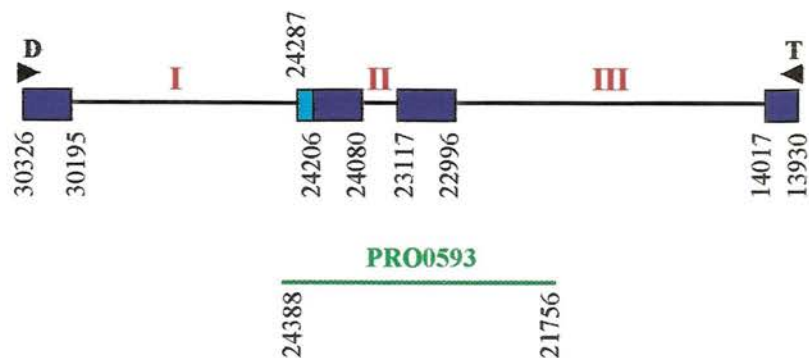


Figura 21: Desenho esquemático da região genômica amplificada pelos iniciadores D e T. Os números representam a posição dos nucleotídeos no clone genômico CTD-2022G18 (AC073184). Os iniciadores D e T estão indicados pelas flechas. Os quadrados azuis indicam os éxons e o quadrado azul claro representa os 81nt presentes em apenas uma das duas isoformas de HIPK2 descritas até o momento. A linha preta representa os íntrons que são nomeados pelos números romanos em rosa. Em verde está representado o mRNA de PRO0593.

Fizemos uma busca por ESTs humanas nestes três íntrons. O primeiro possui 6 ESTs (figura 22A), duas, que alinham na mesma região do clone genômico, provêm de uma mesma biblioteca derivada de tumor de paratireóide (número de acesso no *GenBank*: AI025843 e W38020), uma ORESTES derivada de tumor de sistema nervoso (BI040743), duas alinhadas em outra região do clone genômico derivadas respectivamente de células B de leucemias linfocíticas e de tumor de Schwannoma (BE675563 e AA578866) e uma última derivada de rim (BG400778). A partir destas 5 primeiras ESTs desenhamos iniciadores *forward* e *reverse* para tentar amplificar a possível terceira isoforma.

O segundo íntron desta região não apresentou ESTs humanas e o terceiro apresentou duas ESTs em regiões diferentes (figura 22B), sendo que uma (AI064810) é similar ao mRNA que codifica a proteína PRO0593 (AF111850). PRO0593, de função desconhecida, é codificada por parte da fita reversa da que codifica HIPK2 (figura 21).

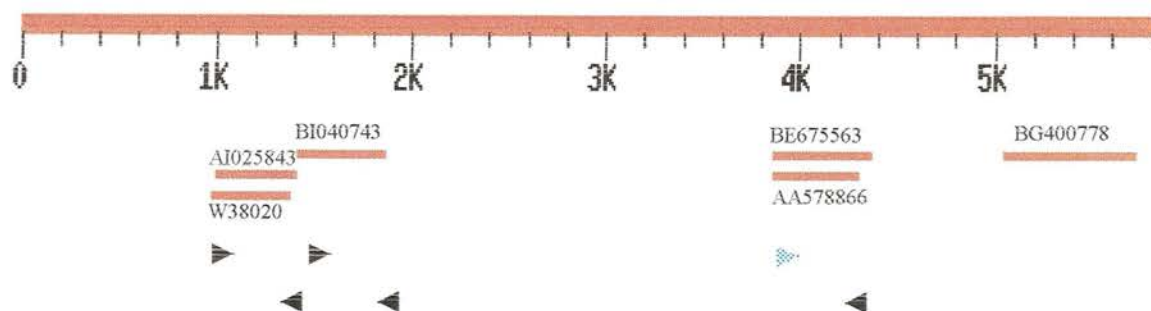
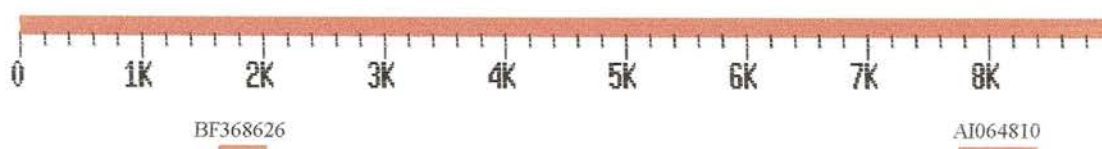
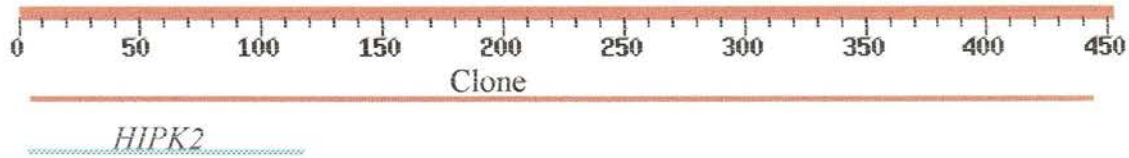
A**B**

Figura 22: Desenho esquemático das ESTs humanas similares a regiões dos introns I (A) e III (B). A linha vermelha mais grossa representa os introns e os números o tamanho da sequência em kilobases. As linhas menores representam ESTs nomeadas pelos seus números de acesso no *GenBank*. As setas em A simbolizam os iniciadores desenhados a partir das sequências de ESTs localizadas acima das setas. A seta azul representa o único iniciador que funcionou nesta região.

Foram feitas RT-PCR a partir de cDNAs derivados de mRNA da linhagem celular de gliomas humanos A172 e de tumor de tireóide utilizando uma combinação de iniciadores novos com os iniciadores D e T. Apenas a dupla TI3 e D funcionou, resultando uma nova EST de aproximadamente 450pb com aproximadamente 110nt similares ao transcrito de *HIPK2*, porém toda similar ao clone humano (figura 23). Buscas no banco de ESTs humanas mostraram 100% de similaridade entre 328nt de 508nt da EST de número de acesso BE675563 e 99% de similaridade entre 293nt de 435nt da EST AA578866. Análises desta EST mostrou que o iniciador reverse de TI3 não funcionou pois a região da EST BE675563 que o desenhemos não se encontra nesta possível nova isoforma.

A



B

> AC073184 Homo sapiens BAC clone CTD-2022G18 from 7, complete sequence
Length = 71607

Identities = 328/328 (100%)
Strand = Plus / Minus

Query: 117 caggtatcaaaggcatctgtgaaacaggaggaaggagcaccctccaatgcaggcgtcat 176
|||||
Sbjct: 28607 caggtatcaaaggcatctgtgaaacaggaggaaggagcaccctccaatgcaggcgtcat 28548

Query: 177 gtcaagtggagtctccccagcaggctgtccttcccatggatgttttctgccaggctgga 236
|||||
Sbjct: 28547 gtcaagtggagtctccccagcaggctgtccttcccatggatgttttctgccaggctgga 28488

Query: 237 gaaactgtggtccctgaccagacgcgggggtggggggccaccgccatcttgcgtggtgatatg 296
|||||
Sbjct: 28487 gaaactgtggtccctgaccagacgcgggggtggggggccaccgccatcttgcgtggtgatatg 28428

Query: 297 acagccagcatatccactccaaagggtgtcgcaaagactctgcaaaactaaaagggttgc 356
|||||
Sbjct: 28427 acagccagcatatccactccaaagggtgtcgcaaagactctgcaaaactaaaagggttgc 28368

Query: 357 tcatcaaaaggcagttgtcaataaagaacaccogtcttctccacggcactgagaccatcc 416
Sbjct: 28367 tcatcaaaaggcagttgtcaataaagaacaccogtcttctccacggcactgagaccatcc 28308

Query: 417 agatctggcttgggactggggtgaagta 444
|||||
Sbjct: 28307 agatctggcttgggactggggtgaagta 28280

Identities = 116/117 (99%), Gaps = 1/117 (0%)
Strand = Plus / Minus

Query: 6 tgtatgacacgngtgaaccagagcaaaaccctttcatcacgcacgtggccccagcacg 65
|||||
Sbjct: 30307 tgtatgacacgg-tgaaccagagcaaaaccctttcatcacgcacgtggccccagcacg 30249

Query: 66 tccaccaacctgaccatgacctttaacaaccagctgaccactgtccacaaccaggta 122
|||||
Sbjct: 30248 tccaccaacctgaccatgacctttaacaaccagctgaccactgtccacaaccaggta 30192

```

> AF326592 Homo sapiens homeodomain interacting protein kinase 2 (HIPK2) mRNA,
    complete cds
    Length = 3984

Identities = 113/114 (99%), Caps = 1/114 (0%)
Strand = Plus / Plus

Query: 6      tgtatgacacggntgaaccagagcaaaacccctttcatcacgcacgtggccccagcag 65
            |||
Sbjct: 1825  tgtatgacacgg-tgaaccagagcaaaacccctttcatcacgcacgtggccccagcag 1883

Query: 66      tccaccaacctgaccatgacctttaacaaccagctgaccactgtccacaaccag 119
            |||
Sbjct: 1884  tccaccaacctgaccatgacctttaacaaccagctgaccactgtccacaaccag 1937

```

Figura 23: *Análise do seqüenciamento do produto da amplificação com os iniciadores D e TI3.* Em A esta representada a seqüência da EST obtida do produto da RT-PCR (linha vermelha mais grossa), a seqüência do clone genômico (linha vermelha menor) que possui homologia com este produto e a seqüência de *HIPK2* (linha verde). Os números indicam o tamanho da EST em pares de base. O resultado das análises de BLAST está indicado em B. Os números indicam a posição da EST (Query) e do gene *HIPK2* (Sbjct).

5. Discussão

5. 1. Obtenção da sequência completa do cDNA do gene humano HIPK2 e análises das sequências de nucleotídeos e aminoácidos

5.1a. Obtenção da sequência completa do cDNA do gene humano HIPK2

Para a obtenção da sequência completa de *HIPK2* utilizou-se o seqüenciamento de clones do IMAGE, RT-PCR e *Shot Gun* PCR . A técnica de *Shot Gun* PCR consiste em uma PCR utilizando iniciadores degenerados com uma primeira ciclagem de baixa estringência e outra de alta estringência para obter-se bandas de diversas regiões da sequência molde (artigo em preparação por nosso grupo). Esta técnica se mostrou bastante eficaz e rápida para obtenção de novas sequências derivadas de regiões internas do gene alvo.

Inicialmente testou-se a possibilidade de varredura de bibliotecas de tecidos de cérebro fetal, tumor de mama, tumor de testículo e tumor de cólon utilizando a técnica de PCR. Porém nenhuma biblioteca apresentou amplificação. Provavelmente não foi possível a amplificação deste gene a partir destas bibliotecas por sua pequena expressão, confirmada por *Northern Blot*, onde só foi possível visualizar *HIPK2* quando utilizadas membranas de mRNA. Dados de *Dot Blot* mostraram uma maior expressão em diversas regiões do cérebro adulto, por isto usou-se RNA de lobo frontal para obtenção do cDNA completo de *HIPK2*.

Observando a sequência da proteína codificada pelo cDNA aqui descrito vimos a presença de 2 metioninas no início da sequência de aa, separadas por 7 aa (figura 3). Comparando com a proteína descrita em camundongo observamos que esta foi descrita começando na segunda metionina, apesar do cDNA codificar as duas. A justificativa é que

a sequência que codifica a segunda metionina apresenta duas características importantes para a escolha do sítio de iniciação da tradução descritas por Kozak, 1986. A primeira é a presença de uma purina 3 nucleotídeos antes do sítio AUG do mRNA e a segunda é a presença de uma guanina logo após o sítio de iniciação. Porém, apesar de uma pequena quantidade, o início pode ocorrer na primeira metionina. Para não correr o risco de omitir dados mantivemos a primeira metionina, mas acreditamos na iniciação da tradução a partir da segunda.

5.1b. Domínios da Proteína Codificada

Análises da sequência de aa de HIPK2 mostraram identidade com PKM (96%) e Hipk2 (94%). PKM foi identificada pela técnica de *yeast two-hybrid*, como sendo uma proteína que interage com proteínas da família Mx (TROST et al. 2000). Proteínas Mx são guaninas trifosfatases (GTPases) altamente reguladas por interferons tipo I e possuem uma atividade antiviral contra vários retrovírus (TROST et al. 2000). Porém testes de imunoprecipitação e *pull-down assays* não confirmaram esta interação (TROST et al. 2000), podendo esta ser uma interação transitória, fraca, ou até mesmo inespecífica. Visto o alto percentual de identidade de PKM com Hipk2 (98%), esta deve ser um ortólogo em hamster.

HIPK2 possui um domínio de quinase com 98% de identidade com Hipk2. Aparentemente este domínio foi altamente conservado durante a evolução, sendo encontrado em *C. elegans*, drófila, levedura e mamíferos, inclusive humanos na proteína PKY (figura 4), uma serina-treonina quinase altamente expressa em fenômenos de multi-resistência a drogas encontrados no tratamento de câncer com quimioterápicos (BEGLEY et al. 1997). Este domínio quinásico parece ter evoluído a partir do domínio encontrado na proteína YAK de levedura, que está envolvida na inibição do crescimento celular

(HARTLEY et al. 1994). Pierantoni e colaboradores (2001) recentemente identificaram esta mesma atividade de inibição do crescimento em HIPK2.

O domínio quinase de HIPK2, além de possuir alta identidade com o domínio de Hipk2 (98%) e PKM (99%), ambas proteínas com comprovada atividade de serina-treonina quinases, possui, assim com PKY, duas seqüências de aminoácidos características deste tipo de quinase, descrita por Begley e colaboradores em 1997, sendo assim HIPK2 também uma serina-treonina quinase. O programa InterPro (www.ebi.ac.uk/interpro/scan.html), um banco de dados de domínios de proteínas, também considerou esta região como um domínio de serina-treonina quinase ativo.

Analizamos também outros domínios de HIPK2, caracterizados primeiramente em Hipk2 (figura 5). Primeiramente comparou-se a seqüência de aa de HID entre a proteína humana e a de camundongo, onde verificamos a ausência de 27 aa no início do possível domínio de HID de HIPK2 e uma identidade do restante de 97%. Análise desta região genômica mostrou a presença de um íntron antes do início do *gap* encontrado na proteína humana com dois sítios *acceptors* (figura 9), o primeiro dando origem ao *gap* e o segundo 81pb adiante dando origem aos 27 aa com 77% de identidade e 91% de positividade com os 27aa encontrados antes apenas na proteína do camundongo. Em março de 2001 foi depositada no *GenBank* uma isoforma do camundongo idêntica a nossa (AF208292) e uma isoforma humana (HOFMANN et al. 2000) similar a primeiramente descrita em murinos (AF208291). Fizemos uma análise *in silico* do número de ESTs humanas, representando estas duas isoformas no *GenBank* que mostrou cinco ESTs nesta região, das quais uma (N43800), derivada de uma biblioteca de melanócitos normais, possuía estes 81pb a mais e as outras, todas ORESTES, derivadas de medula óssea normal, tumor de sistema nervoso e duas de linhagem celular de cólon tumoral, confirmavam o transcrito aqui descrito (BF904758, BF935452, BF814194, BF814195).

O domínio co-repressor de Hipk2 é responsável, em conjunto com o HID, pelo aumento da atividade repressora de NK (KIM et al. 1998). Este tem 100% de identidade entre as duas proteínas, humana e murina (fig 5), sugerindo uma conservação desta atividade repressora na proteína humana.

Estudos com Hipk2 e PKM demonstraram serem estas proteínas nucleares localizadas nos *nuclear speckles* ou *nuclear dots* (KIM et al. 1998, 1999; TROST et al. 2000), estruturas nucleares de função desconhecida. Estudos de deleções em Hipk2 mostraram dois sinais de localização nuclear (NLS) e um sinal de retenção no *nuclear speckle* (SRS) (KIM et al. 1998, 1999). HIPK2 possui uma alta identidade com estes sinais - 100% com NLS1, 98% com NLS2 e 94% com SRS, como visto na figura 5.

Estudos com a proteína Hipk2 mostraram sua interação com a proteína SUMO-1 (KIM et al. 1999). SUMO-1 faz parte de uma família de proteínas semelhantes às ubiquitinas (KIM et al. 1999). Diferentemente das ubiquitinas, envolvidas na degradação seletiva de algumas proteínas, SUMO-1 parece estar envolvida com a localização subcelular ou aumento da estabilidade das suas proteínas alvo (KIM et al. 1999). Modificações feitas por SUMO-1 em HIPK2 estão relacionadas com a sua localização nos *nuclear dots*. A associação de SUMO-1 com Hipk2 ocorre por uma interação com mUBC9, uma enzima (E2) de conjugação de Ubl, que interage com Hipk2 através de uma sequência mínima de aa do domínio SRS (aa 860 a 894 de Hipk2) que possui 100% de identidade com HIPK2. Experimentos para comprovar esta interação de HIPK2 com SUMO-1 através de uma enzima de conjugação de Ubl serão necessários para provar que esta mesma interação ocorre no homem.

A sequência PEST, prolina (P), ácido glutâmico (E), serina (S) e treonina (T), é encontrada em HIPK2 e possui 100% de identidade com a proteína de camundongo. Uma hipótese para a sequência PEST, sugerida por Rogers e colaboradores, em 1986, diz que

esta sequência de aa faz com que a proteína seja mais suscetível a uma rápida proteólise, podendo esta ser impedida por alguns fatores como a localização intracelular ou a associação com outras proteínas. No caso de Hipk2 e talvez de HIPK2 a sua localização nuclear, devida provavelmente a alterações feitas por SUMO-1, possivelmente evita esta degradação. Recentemente foi descrito por Pierantoni e colaboradores (2001) uma interação de HIPK2 com a família de HMGI(Y), *High Mobility Group I*, que são proteínas arquiteturais das cromatinas que parecem afetar a transcrição e estão altamente expressas no desenvolvimento. Esta interação ocorre através o domínio PEST é direta e específica (PIERANTONI et al. 2001). Estudos estão sendo realizados para identificar a função desta família HMGI(Y) e a função da ligação a HIPK2. Pierantoni e colaboradores, 2001, sugerem que em conjunto com outras proteínas HIPK2 e HMGI(Y) seriam responsáveis pela a formação de um complexo regulador da transcrição.

Um último domínio encontrado em Hipk2, localizado no C-terminal desta proteína, é o domínio YH, uma região enriquecida com resíduos de histidina e tirosina, de função desconhecida e conservado entre as HIPKs. A proteína humana aqui descrita apresenta 95% de identidade com este domínio. Este domínio YH é encontrado em todas as proteínas da família HIPK (com 58% de identidade com Hipk1 e 50% com Hipk3 e HIPK3) e também na proteína PKY, com 50% de identidade. Neste domínio foi encontrada uma variação entre a sequência de Hipk2 e HIPK2. Esta consiste em uma região repetitiva de alaninas que é característico apenas nas proteínas HIPK2, não sendo encontrado nas outras proteínas.

É interessante comentar aqui as diferentes possíveis atividades de HIPK2 descritas recentemente, apesar de não terem domínios específicos caracterizados. Como já comentado anteriormente, HIPK2 é membro de uma família HIPK que se liga a família de genes homeobox NK e aumenta a atividade repressora destes. Recentemente Wang e

colaboradores, 2001, mostraram que HIPK2 ativa a transcrição de genes regulados por p53 via a inibição da degradação desta proteína. Esta inibição é feita pela regulação feita por HIPK2 na atividade de Mdm2 provavelmente via fosforilação de Mdm2, impedindo assim que esta promova a degradação de p53. Pierantoni e colaboradores mostraram que HIPK2 tem uma atividade de inibição do crescimento celular, podendo ser um complemento da atividade de ativação de p53 que para a divisão celular para verificar irregularidades nas divisões e não obrigatoriamente leva a apoptose. Li e colaboradores, 2000, mostraram a interação de HIPK2 com TRADD (*TNF-R1 associated death domain*) que está envolvido na sinalização mediada por TNF. Porém esta ligação é inibida pela presença de receptores de fatores de morte (*death factors*) e os autores não concluem a função desta possível interação.

5.1c. Análises Genômicas

As análises genômicas indicam para a existência de possivelmente 16 éxons, visto que esta região não está totalmente seqüenciada. Análises in silico apontam para a existência de apenas uma cópia deste gene no genoma e esta está localizada no cromossomo 7q32-34. O gene de camundongo está localizado no cromossomo 6.

5.1d. Avaliação da variação de alaninas no domínio YH em diferentes espécies

O domínio YH apresentou uma variação no número de alaninas entre o humano e os roedores. No intuito de avaliar uma possível relação filogenética nesta variação analisamos esta região em alguns animais. O padrão de repetição das alaninas não obedeceu o esperado para as diferentes espécies de mamíferos estudadas. É possível concluir que este domínio é altamente preservado entre as diferentes espécies, visto a similaridade

encontrada no restante desta região, sendo possivelmente fundamental para a função da proteína.

5.2. *Análise da expressão deste gene*

5.2a. *Northern Blot*

Não se obteve resultados de hibridizações de membranas de RNA total, provavelmente devido a baixa expressão deste gene, visto que a sonda funcionou em membranas de mRNA. Análises de *Northern Blot* a partir de mRNA (fig.13) indicam a presença de pelo menos três transcritos de mais de 10kb, aproximadamente 7kb e aproximadamente 4kb, correspondendo à isoforma aqui descrita e à descrita em março, visto a pequena diferença de tamanho entre estas duas isoformas.

5.2b. *Membrana de Dot Blot*

Os resultados da hibridização do *Dot Blot* (fig.14a) indicam que este gene é bastante expresso no cérebro adulto, principalmente nos núcleos da base (gráfico 1), coração e rim. São os núcleos da base: núcleo caudado, núcleo lentiforme (dividido em putâmen e globo pálido), amígdala, substância cinzenta ou claustrum, núcleo acubens, núcleo basal de Meynert e subtálamo (MACHADO 1993). Estes núcleos estão envolvidos essencialmente no controle motor e coordenação, principalmente os núcleos caudado e putâmen (MACHADO 1993) e estão relacionados a doenças que afetam os movimentos, como a doença de Parkinson, que aparece geralmente após os 50 anos e é caracterizada por tremores de repouso em extremidades, rigidez da musculatura esquelética, lentidão e redução da atividade motora espontânea e dificuldade protocinética, e a doença de Huntington, que aparece em indivíduos entre 20 e 50 anos e se caracteriza por coréia e

demência (MACHADO 1993; GUYTON e HALL 1996). O núcleo basal de Meynert, também conhecido por *substância inominata*, é relacionado com a memória e com funções psíquicas superiores e aumenta progressivamente de tamanho conforme aumenta a escala filogenética e está envolvida na doença de Alzheimer (MACHADO 1993). Estudos mais aprofundados buscando uma possível relação entre este gene e estas doenças podem ser de grande importância.

O gráfico 2 mostra a expressão deste gene em diferentes regiões do coração e da aorta, mostrando maior expressão no ventrículo esquerdo e no septo interventricular, que é formado principalmente pelo ventrículo esquerdo. O gráfico 4 compara o tecido adulto com o tecido fetal e linhagens tumorais, quando existentes. É importante notar que a expressão de *HIPK2* é maior no cérebro adulto que no fetal, podendo-se inferir uma função de conservação e manutenção do órgão adulto. O mesmo ocorre no coração, fígado e pulmão. Este gene deve possuir diferentes funções, sendo uma delas já descrita como um co-repressor de certos grupos de genes *homeobox*, envolvidos no desenvolvimento, sendo esperada sua maior expressão no tecido adulto, onde já houve o desenvolvimento do órgão, não sendo mais necessária a expressão do grupo de genes *homeobox*. Em rim e baço sua expressão é relativamente constante, tanto em tecidos adultos como fetais, mostrando uma possível função na diferenciação durante a formação dos órgãos e manutenção deste fenótipo diferenciado.

A amostra de pulmão é a mais completa (gráfico 4), sendo composta de tecido adulto, fetal e uma linhagem tumoral. O tecido fetal e a linhagem tumoral têm menor expressão respectivamente que o tecido adulto, um padrão comum no câncer, onde a expressão de alguns genes reguladores volta a ter padrões similares aos observados nos estágios embrionários e fetal, onde as células estão indiferenciadas e com alto índice mitótico.

Comparando-se as linhagens celulares de diferentes leucemias com os tecidos ou células normais mais próximos da origem destas linhagens (gráfico 3) notou-se menor expressão deste gene nas leucemias. Esta comparação é de certo modo grosseira, visto que estamos comparando linhagens celulares com tecidos ou células normais. Esta comparação está sendo usada como uma indicação da possível alteração da expressão deste gene em leucemias, sendo necessária sua comprovação em tecidos normais e tumorais. *HIPK2* pode estar envolvido na gênese das leucemias visto que é um regulador de genes *homeobox*, que já foram amplamente relacionados com este tipo de câncer (MOSKOW et al. 1995).

Apesar de não comprovado, é possível que este gene esteja envolvido na carcinogênese de alguns tecidos. Estudos de expressão (como RT-PCR), de mutações e de alterações na expressão deste gene por animais transgênicos e *knock-out* seriam de grande importância para conhecer melhor sua função e seu envolvimento com o câncer.

5.2c. Análises de SAGE (*Serial Analysis of Gene Expression*)

A análise de SAGE do transcrito de *HIPK2* mostrou uma expressão em poucos tecidos, visto que de 100 bibliotecas, apenas 23 apresentavam *tags* para este transcrito. Porém, as análises de *Dot Blot* mostraram uma expressão ampla, apesar de pequena deste gene. Isto pode ser explicado pelo fato de que para genes de baixa expressão são necessários seqüenciamentos de vários *tags* para aumentar a possibilidade de obtenção dos *tags* mais raros dentro de uma biblioteca. A análise comparativa do número máximo de *tags* obtidos de *HIPK2* (277), do transcrito que codifica *p53*, de expressão moderada (370) e *GAPDH* (*glyceraldehyde-3-phosphate dehydrogenase*), considerado um gene constitutivo (6958) confirma a baixa expressão de *HIPK2*. Esta observação é confirmada com a dificuldade de se obter resultados nas membranas de *Northern Blot* feitas a partir de RNA total. A maioria das bibliotecas encontradas foi derivada do cérebro (10/23), coerente

com sua alta expressão neste órgão, observada nas análises de *Dot Blot*. Na membrana de *Dot Blot* o ovário aparentava uma pequena expressão, no entanto, as análises de SAGE indicam a maior expressão deste em uma linhagem celular tumoral de ovário, o que poderia indicar um papel deste gene na tumorigênese de ovário.

Análises do número de ESTs encontradas no UNIGENE mostraram a existência de 94 ESTs do transcrito de *HIPK2* (Hs.236131). A comparação novamente com os transcritos que codificam p53, que possui 241 ESTs (Hs.1846) e GAPDH, que possui 11.596 ESTs (Hs.169476), ressalta a baixa expressão deste gene, já indicada por *Northern Blot* e SAGE.

5.2d. Hibridização *in situ* de *HIPK2* em camundongo fetal

HIPK2 é um co-repressor da família NK de genes homeóticos. Esta família é classificada como genes órfãos, que possuem maior importância na organogênese, ao contrário dos Hox, ou genes agrupados, que têm função principal na formação do molde corporal (MARK et al. 1997). Isto pode justificar uma expressão mais específica em alguns órgãos, como foi observado em nossos resultados, e não uma expressão difusa pelos cortes fetais (figura 15). É interessante notar a sua expressão precisa no tecido epitelial (figura 15 B, C e E) e talvez relacioná-la com a interação epitélio-mesentério, que ocorre na formação da maioria dos órgãos (BIRCHMEIER e BIRCHMEIER 1993). Esta interação é de importância fundamental para o desenvolvimento e a regeneração dos tecidos, tendo uma significância clínica na etiologia dos tumores (BIRCHMEIER e BIRCHMEIER 1993). Na morfogênese do pulmão já foi descrita a participação de NKX2.1 na interação epitélio-mesênquima (COSTA et al. 2001). *HIPK2* apresentou sinais de hibridização nos brônquios em desenvolvimento (figura 15E), podendo estar agindo sobre NKX2.1.

O fato de não se ter encontrado sinal de hibridização no coração (órgão onde já é conhecida a influência da família NK no seu desenvolvimento) de fetos com 13dpc pode

ser justificado pela expressão diferencial e temporal dos genes homeóticos no desenvolvimento cardíaco. Para uma melhor visualização e entendimento da função de *HIPK2* no desenvolvimento são necessárias hibridizações sucessivas dos diferentes estágios do desenvolvimento fetal do camundongo.

O tubo neural e a notocorda apresentaram marcação (fig.9A) e a morfogênese do tubo parece depender da interação com a notocorda assim como de outros tecidos (BRONNER-FRASER e FRASER 1997; TEILLET et al. 1998). Estes resultados frisam a importância de *HIPK2* no sistema nervoso.

5.2e. RT-PCR de HIPK2 em camundongos

As diversas RT-PCR para avaliar a expressão das diferentes isoformas de *HIPK2* foram realizadas utilizando os iniciadores D e T e estes resultam na amplificação de três isoformas, de diferentes tamanhos. O uso de um mesmo par de iniciadores para a amplificação das três bandas distintas faz com que esta PCR seja competitiva. Neste caso, supondo um mesmo nível de expressão das diferentes isoformas, teríamos um gradiente de amplificação inversamente proporcional ao tamanho das bandas. Qualquer alteração deste gradiente indica uma modificação no nível de expressão das diferentes isoformas. A isoforma com menor tamanho é amplificada com maior eficiência, pois há maior probabilidade da polimerase copiar toda a molécula de DNA a cada ciclo de amplificação. Deste modo, os fragmentos menores aumentam exponencialmente, enquanto os fragmentos maiores têm um crescimento menos eficiente do seu número de moléculas. Vários trabalhos já utilizaram esta metodologia (THELEN et al. 1999; HU et al. 2000, VOGIAGIS et al. 2000).

I. Análises das isoformas de HID no desenvolvimento do camundongo

Os dados obtidos na RT-PCR dos diferentes estágios do desenvolvimento do rim, cérebro e coração conduzem à hipótese de que a isoforma que descrevemos, que não codifica os primeiros 27 aminoácidos (aa) do HID, esteja competindo com o domínio de ligação. Porém pela ausência destes primeiros aa é possível que a interação com o homeodomínio de HIPK2 seja diferente entre as isoformas e este processo seja importante e regulatório. Experimentos complementares serão necessários para validar ou não esta hipótese. Anteriormente sugerimos que a principal função do gene *HIPK2* era na manutenção do órgão adulto. Este experimento e a hibridização *in situ*, anteriormente comentada, indicam que este gene também é importante no desenvolvimento.

Os dados obtidos pelas RT-PCR dos diferentes estágios do desenvolvimento mostram que a isoforma menor é prevalente em tecidos adultos e a isoforma contendo os 81nt tem sua expressão variada nos diferentes tecidos. Em rins estas isoformas têm a sua expressão invertida nos estágios fetais e adulto. Porém para obtermos conclusões mais significativas precisaremos obter uma maior amostra e um maior número de estágios, em conjunto com uma avaliação precisa da intensidade das bandas por densitometria e testes estatísticos das variações encontradas.

II. Análise das isoformas de HID em tecidos e linhagens celulares humanos

Devido ao único par de amostras de tecidos normais e tumorais, não é possível inferir sobre o papel deste gene e suas isoformas na carcinogênese, havendo apenas uma indicação que no tumor de língua há uma inversão da expressão das duas isoformas. Porém, este experimento confirma a existência de uma terceira isoforma de *HIPK2* que será posteriormente comentada.

III. Análise das isoformas de HIPK2 em linhagens celulares de gliomas de humanos e ratos

Para melhor avaliar a expressão das diferentes isoformas de *HIPK2* em tumores analisamos amostras de linhagens celulares de gliomas humanos tratadas e não tratadas com ácido retinóico por 48h (figura 18). A reação desta droga nas linhagens de gliomas humanos já foi estudada pelo grupo de pesquisa da Profa. Sogayar e sabe-se que provoca uma diferenciação das linhagens tumorais, e que em até 48h esta diferenciação já ocorreu.

As duas linhagens de gliomas apresentaram variações diferentes da expressão das duas isoformas. Como estas linhagens ainda estão em processo de caracterização em relação à malignidade e resposta ao ácido retinóico, não é possível interpretar este diferente padrão de expressão das duas isoformas nas duas linhagens após o tratamento com ácido retinóico. Seria interessante avaliar a expressão destas três bandas em diferentes tecidos, normais e tumorais.

Para caracterizar a expressão das diferentes isoformas de *HIPK2* o laboratório da Profa. Sogayar nos forneceu amostras de RNA de duas variantes da linhagem de glioma de ratos, C6, a variante ST1 e a P7. Verificou-se que a hidrocortisona inibe o crescimento da variante celular ST1 e restaura reversivelmente o fenótipo normal destas células (ARMELIN et al. 1983). Em contrapartida, a variante celular P7 é resistente ao tratamento com esta droga. A variante ST1, após aproximadamente cinco horas de tratamento começa a apresentar alterações moleculares, como a parada de crescimento e após aproximadamente 24 horas estas células reverterem morfológicamente.

Estes resultados nos mostram primeiramente que a variante sensível ao tratamento tem um aumento da expressão de *HIPK2* no decorrer do tratamento. Este efeito é principalmente observado na segunda isoforma, atingindo o auge com 24 horas de

tratamento, quando ocorre a diferenciação morfológica da célula. Após 48 horas de tratamento, o nível de expressão de *HIPK2* tende a se igualar com o controle, onde podemos hipotetizar que as células não respondem mais ao tratamento, voltando à expressão basal de *HIPK2*. O nível de expressão basal deste gene nos mostra que antes do tratamento a variante sensível à hidrocortisona já apresentava a segunda banda com maior intensidade que a primeira. Já na variante P7, esta expressão basal tem a primeira isoforma mais intensa. Durante seu tratamento, há uma diminuição contínua da expressão de *HIPK2* e uma tendência das duas bandas igualarem sua intensidade. Apenas após 48 horas de tratamento observamos maior intensidade na segunda banda. Em resumo, *HIPK2* parece ser ativado pelo tratamento na variante sensível, porém em P7 parece haver uma supressão de sua ativação. Sendo este gene responsável pela manutenção da estrutura do tecido normal, podemos cogitar que sua inativação parece tornar as células mais indiferenciadas, tornando-as refratárias ao tratamento. A expressão da segunda isoforma, com 550pb, parece ser importante para esta diferenciação. Para melhor validar este experimento será necessária a quantificação das intensidades das bandas por densitometria e testes estatísticos das variações encontradas.

5.3. Análises da possível terceira isoforma de *HIPK2*

Os resultados obtidos nas RT-PCRs acima citadas nos indicavam a presença de uma possível nova isoforma na região do *HID*, maior que as já citadas. A procura de ESTs humanas localizadas nos três íntrons desta região amplificada pelos iniciadores D e T mostraram que o íntron aqui nomeado I (figura 21) apresentou algumas poucas regiões representadas por ESTs, uma destas regiões era representada por duas ESTs de bibliotecas diferentes, aumentando a chance de não ser uma contaminação. Elegemos este íntron como

possível candidato a possuir um éxon ainda não descrito. Desenhamos iniciadores *forward* e *reverse* baseados nas ESTs de três regiões que acreditamos serem candidatas a codificarem esta nova isoforma (figura 22). Apenas um iniciador funcionou nas reações de RT-PCR. Análises deste produto indicam a presença de um novo éxon que codifica a terceira isoforma de *HIPK2* na região de HID. Análises mais detalhadas e completas são necessárias para a descrição desta nova isoforma.

6. Conclusão

Este trabalho visou inicialmente a caracterização de um novo gene humano que havia anteriormente apenas sido investigado em camundongos e hamsters. Após a obtenção completa da sequência de nucleotídeos do cDNA de *HIPK2*, derivamos a sequência de aa e com esta analisamos vários domínios da nova proteína, comparando-os com os previamente descritos na proteína de camundongo. Nesta etapa observamos que a proteína *HIPK2* é bastante conservada entre roedores e humanos. Seu domínio quinase pertence a uma nova família de serina-treonina quinases derivadas da proteína Yak1p de leveduras, onde esta proteína tem importante função na regulação do ciclo celular.

As análises da região polimórfica do domínio YH, apesar de não apresentarem uma relação filogenética, mostram que esta proteína está presente em outros vertebrados e que o domínio YH é bastante conservado, enquanto que o número de repetições de alaninas é bastante variável entre os diferentes mamíferos estudados. Este alto nível de conservação sugere um importante papel deste domínio (de função ainda não conhecida).

A impossibilidade de obter-se sinais de hibridização nas membranas constituídas com RNA total e a necessidade de um longo tempo de exposição das membranas contra filmes de raio X ou nos cassetes especiais para o aparelho Storm sugerem a baixa expressão deste gene de modo geral. Estas análises podem ser confirmadas pela análise de SAGE e UNIGENE, que quando comparadas com análises de genes de expressão moderada e genes constitutivos é pequena.

As análises de *Dot Blot* mostraram maior expressão deste gene em tecidos adultos. A partir deste resultado podemos inferir que, como se trata de um gene repressor de genes homeóticos, sua expressão durante o desenvolvimento é variada e tende a ser menor que no adulto, já que durante o desenvolvimento os genes *homeobox* estão altamente ativos para a

formação dos diversos tecidos. Sendo assim seria esperado o encontro aumentado do nível de expressão de *HIPK2* em órgãos já desenvolvidos, reprimindo a atividade de certos genes homeóticos não mais necessários. Este pressuposto não impede a sua expressão no desenvolvimento, nem inferioriza sua importância neste período, pois apesar de pouco expresso, ele é ativo. Dados da hibridização *in situ* confirmam este fato e mostram por sua expressão específica nos tecidos a importância deste gene no desenvolvimento. Relacionado as hipóteses sugeridas acima, podemos também inferir outras funções de *HIPK2* não relacionadas à sua atividade repressora. Isto pode ser visto pela recente publicação que a proteína HIPK2 ativa a transcrição mediada pela proteína supressora de tumor p53 (WANG et al. 2001). Por ser uma nova proteína, há ainda poucos estudos sobre sua função. Dados de *Dot Blot* nos levam a inferir um papel na manutenção dos órgãos adultos, principalmente o cérebro, coração e fígado.

É interessante notar que no rim há uma elevada expressão deste gene tanto em órgãos adultos como fetais e os dados de hibridização *in situ* também mostraram expressão nos precursores deste órgão. A soma destes dados permite sugerir um papel constitutivo deste gene em rins.

Dados de *Dot Blot* também mostraram um alto nível de expressão de *HIPK2* em cérebro, principalmente em alguns núcleos da base. Estes núcleos são responsáveis pelo controle motor e de coordenação e estão envolvidos em doenças como Parkinson e Huntington. Análises de SAGE mostraram uma prevalência deste transcrito em bibliotecas derivadas de cérebro e os resultados da hibridização *in situ* indicam uma importância deste gene no desenvolvimento do sistema nervoso. A somatória destes dados indica uma alta importância deste gene neste órgão. Genes da família NK já são conhecidos por sua importância na formação e manutenção de diversos constituintes do sistema nervoso.

Dados da hibridização *in situ* mostram uma expressão específica nos tecidos epiteliais. Estes dados apontam para uma possível participação deste gene nas interações epitélio-mesentério durante o desenvolvimento. Esta interação envolve o crescimento e modelamento epitelial em resposta a fatores parácrinos mesenquimais, como liberação de fatores de crescimento e resposta de fatores de transcrição levando a diferenciação através da alteração na expressão gênica (BIRCHMEIER e BIRCHMEIER 1993) e já descrita em diversos órgãos como glândulas mamárias, próstata, rim e pulmão (CUNHA 1994; PRICE e STILES 1996; ALCORN et al. 1999; BALKOVETZ e LIPSCHUTZ 1999; BURROW 2000; COSTA et al. 2001).

De modo geral este gene se mostrou, por avaliações de sua expressão, importante na formação e manutenção do cérebro e rim. Sua expressão no coração e dados da literatura que apontam a importância dos genes da família NK, principalmente *Nkx2.5*, na morfogênese do coração também indicam a importância da expressão de *HIPK2* no coração. Este órgão ilustra bem o possível papel repressor, visto que ele é menos expresso no tecido fetal, onde os genes *homeobox* são bastante ativos, e tem alta expressão no tecido adulto, onde há a diminuição da expressão desta família NK.

Os estudos de RT-PCR mostram uma terceira isoforma deste gene no domínio HID. Esta isoforma possui aproximadamente 600pb a mais do que a aqui descrita. A amplificação desta terceira isoforma é realizada por uma PCR competitiva, visto que a mesma ocorre simultaneamente com outras duas de tamanho bem mais reduzido. Por se tratar deste tipo de PCR, a quantidade de DNA obtido desta isoforma não é suficiente para isolá-la a partir de um gel de agarose. Análises de ESTs e a amplificação de parte desta isoforma indicam sua real existência.

Neste trabalho caracterizamos a sequência de um novo gene humano, do qual outra isoforma foi publicada. A diferença destas duas isoformas é de apenas 81nt localizados no

início de HID. Posteriormente visualizamos uma terceira e ainda maior isoforma nesta mesma região. Estas três isoformas podem estar envolvidas em diferentes ligações com as homeoproteínas, indicando possivelmente diferentes funções para HIPK2.

Ao avaliarmos sua expressão, concluímos que *HIPK2* é importante tanto para o desenvolvimento, quanto para a manutenção de alguns órgãos diferenciados, principalmente o cérebro, coração e rim, ressaltando a hipótese de múltiplas funções para esta nova proteína.

7. Perspectivas futuras

Este projeto realizou a primeira etapa da pesquisa pós-genômica, que foi a caracterização de um novo gene humano, avaliação de sua expressão e identificação de novas isoformas. Porém ainda sabe-se pouco sobre a função de *HIPK2*. Pretende-se continuar a estudá-lo na segunda etapa da pesquisa pós-genômica, que é a caracterização de sua função. Esta caracterização será feita por análises *in vivo*. Para isto será construído um camundongo sem um ou dois alelos de *HIPK2* e o fenótipo resultante será avaliado.

Em se tratando de um gene envolvido no desenvolvimento, pretende-se estudar a sua expressão durante este processo por hibridizações *in situ* em camundongos.

FUNDAÇÃO ANTONIO PRUDENTE
ANA MARIA R. A. KUNTINATI
Coordenadora de Pós-graduação

8. Referências Bibliográficas

- Alcorn D, Maric C, McCausland J. Development of the renal interstitium. **Pediatr Nephrol** 1999; 13:347-54.
- Anbazhagan R, Raman V. Homeobox genes: molecular link between congenital anomalies and cancer. **Eur J Cancer** 1997; 33:635-7.
- Armelin MC, Stocco RC, Armelin HA. Control of rat C6 glioma cell proliferation: uncoupling of the inhibitory effects of hydrocortisone hormone in suspension and monolayer cultures. **J Cell Biol** 1983; 97(2):455-8.
- Balkovetz DF, Lipschutz JH. Hepatocyte growth factor and the kidney: it is not just for the liver. **Int Rev Citol** 1999; 186:225-60.
- Begley DA, Berkenpas MB, Sampson KE, Abraham I. Identification and sequence of human Pky a putative kinase with increased expression in multidrug-resistant cells with homology to yeast protein kinase Yak1. **Gene** 1997; 200:36-43.
- Benson DW, Silberbach GM, Kavanaugh-McHugh A, et al. Mutations in the cardiac transcription factor NKX25 affect diverse cardiac developmental pathways. **J Clin Invest** 1999; 104:1567-73.
- Bhatia-Gaur R, Donjacour AA, Sciavolino PJ, et al. Roles for Nkx31 in prostate development and cancer. **Genes Dev** 1999; 13:966-77.
- Biben C, Weber R, Kesteven S, Stanley E, et al. Cardiac septal and valvular dysmorphogenesis in mice heterozygous for mutations in the homeobox gene Nkx2-5. **Circ Res** 2000; 87:888-95.
- Bieberich CJ, Fujita K, He WW, Jay G. Prostate-specific and androgen-dependent expression of a novel homeobox gene. **J Biol Chem** 1996; 271:31779-82.

Birchmeier C, Birchmeier W. Molecular aspects of mesenchymal-epithelial interactions. **Annu Rev Cell Biol** 1993; 9:511-40.

Bowen C, Bubendorf L, Voeller HJ, et al. Loss of NKX31 expression in human prostate cancers correlates with tumor progression. **Cancer Res** 2000; 60:6111-5.

Bronner-Fraser M, Fraser SE. Differentiation of the vertebrate neural tube. **Curr Opin Cell Biol** 1997; 9:885-91.

Burrow C. Regulatory molecules in kidney development. **Pediatr Nephrol** 2000; 14:240-53.

Chan SK, Jaffe L, Capovilla M, Botas J, Mann RS. The DNA binding specificity of ultrabithorax is modulated by cooperative interactions with extradenticle another homeoprotein. **Cell** 1994; 78:603-15.

Church GM, Gilbert W. Genome sequencing. **Proc Natl Acad Sci USA** 1984; 81:1991-5.
Cillo C, Barba P, Freschi G, Bucciarelli G, Magli MC, Boncinelli E. HOX gene expression in normal and neoplastic human kidney. **Int J Cancer** 1992; 51:892-97.

Cillo C, Faiella A, Cantile M, Boncinelli E. Homeobox genes and cancer. **Exp Cell Res** 1999; 248:1-9.

Costa RH, Kalinichenko VV, Lim L. Transcription factors in mouse lung development and function. **Am J Physiol Lung Cell Mol Physiol** 2001; 280:L823-38.

Cunha GR. Role of mesenchymal-epithelial interactions in normal and abnormal development of the mammary gland and prostate. **Cancer** 1994; 74 Suppl 3:1030-44.

De Vita G, Barba P, Odartchenko N, et al. Expression of homeobox-containing genes in primary and metastatic colorectal cancer. **Eur J Cancer** 1993; 29A:887-93.

Dias Neto E, Garcia Correa R, Verjovski-Almeida S, et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence Tags. **Proc Natl Acad Sci U S A** 2000; 97:3491-6.

Doevendans PA, Van Bilsen M. Transcription factors and the cardiac gene programme. **Int J Cell Biol** 1996; 28:387-403.

Edelman GM, Jones FS. Outside and downstream of the homeobox. **J Biol Chem** 1993; 268:20683-6.

Friedmann Y, Daniel CA, Strickland P, Daniel CW. Hox genes in neoplastic mouse mammary gland. **Cancer Res** 1994; 54:5981-5.

Gehring WJ, Affolter M, Burglin T. Homeodomain proteins. **Annu Rev Biochem** 1994; 63:487-526.

Gehring WJ, Qian YQ, Billeter M, et al. Homeodomain DNA-recognition. **Cell** 1994; 78:211-23.

Guyton AC, Hall JE, editors. **Textbook of medical physiology**. 9th ed. Philadelphia: W.B. Saunders; 1996. The cerebellum, the basal ganglia, and overall motor control; p.724-9.

Hartley AD, Ward MP, Garret S. The Yak1 protein kinase of *saccharomyces cerevisiae* moderates thermotolerance and inhibits growth by an Sch9 protein kinase-independent mechanism. **Genetics** 1994; 136:465-74.

Harvey RP. NK-2 homeobox genes and heart development. **Dev Biol** 1996; 178:203-16.

He WW, Sciavolino PJ, Wing J, et al. A novel human prostate-specific androgen-regulated homeobox gene (NKX31) that maps to 8p21 a region frequently deleted in prostate cancer. **Genomics** 1997; 43:69-77.

Hofmann TG, Mincheva A, Lichter P, Droge W, Lienhard SM. Human homeodomain-interacting protein kinase-2 (HIPK2) is a member of the DYRK family of protein kinases and maps to chromosome 7q32-q34. **Biochimie** 2000; 82:1123-7.

Hu Q, Jin LW, Starbuck MY, Martin GM. Broadly altered expression of mRNA isoforms of fe65 a facilitator of beta amyloidogenesis in alzheimer cerebellum and other brain region. **J Neurosci Res** 2000; 60:73-86.

Innis JW. Role of HOX genes in human development. **Curr Opin Pediatr** 1997; 9:617-22.

Kasahara H, Izumo S. Identification of the In vivo casein kinase II phosphorylation site within the homeodomain of cardiac tissue-specifying homeobox gene product Csx/Nkx25. **Mol Cel Biol** 1999; 19:526-36.

Kim Y, Nirenberg M. Drosophila NK-homeobox genes. **Proc Natl Acad Sci USA** 1998; 86:7716-20.

Kim YH, Choi CY, Kim Y. Covalent modification of the homeodomain-interacting protein kinases (HIPK2) by ubiquitin-like protein SUMO-1. **Proc Natl Acad Sci USA** 1999; 96:12350-5.

Kim YH, Choi CY, Lee SJ, Conti MA, Kim Y. Homeodomain-interacting Protein-kinases a novel family of co-repressors for homeodomain transcription factors. **J Biol Chem** 1998; 273:25875-9.

Kozak M. Point mutation define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. **Cell** 1986; 44:283-92.

Krumlauf R. Hox genes in vertebrate development. **Cell** 1994; 78:191-201.

Li X, Wang Y, Debatin K-M, Hug H. The serine/threonine kinase HIPK2 interacts with TRADD but not with CD95 or TNF-R1 in 293T cells. **Biochem Biophys Res Commun** 2000; 277:513-17.

Machado ABM. **Neuroanatomia funcional**. 2. ed. São Paulo: Atheneu; 1993. Estrutura e função dos núcleos da base e do centro branco medular do cérebro; p.249-56.

Mann RS, Chan SK. Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain. **Trends Genet** 1996; 12:258-62.

Mark M, Rijli FM, Chambon P. Homeobox genes in embryogenesis and pathogenesis. **Pediatr Res** 1997; 42:421-9

McGinnis W, Krumlauf R. Homeobox genes and axial patterning **Cell** 1992; 68:283-302.

Moskow JJ, Bullrich F, Huebner K, Daar IO, Buchberg AM. Meis1 a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice. **Mol Cel Biol** 1995; 15:5434-43.

Pabst O, Zweigerdt R, Arnold H. Targeted disruption of the homeobox transcription factor Nkx2-3 in mice results in postnatal lethality and abnormal development of small intestine and spleen. **Development** 1999; 126:2215-25.

Patterson KD, Cleaver O, Gerber WV, Grow MW, Newman CS, Krieg PA. Homeobox Genes in Cardiovascular Development **Curr Top Dev Biol** 1998; 40:1-44.

Petrini M, Quaranta MT, Testa U, et al. Expression of selected human HOX-2 genes in B/T acute lymphoid leukemia and interleukin-2/interleukin-1 beta-stimulated natural killer lymphocytes. **Blood** 1992; 80:185-93.

Pierantoni GM, Fedele M, Pentimalli F. High mobility group I (Y) proteins bind HIPK2, a serine-threonine kinase protein which inhibits cell growth. **Oncogene** 2001; 20:6132-41.

Price WA, Stiles AD. New insights into lung growth and development. **Curr Opin Pediatr** 1996; 8:202-8.

Rogers S, Wells R, Rechsteiner M. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. **Science** 1986; 234:364-8.

Rubenstein JLR, Puelles L. Homeobox gene expression during development of the vertebrate brain. **Curr Top Dev Biol** 1994; 29:1-63.

Sambrook J, Fritsch EF, Maniatis T. **Molecular cloning a laboratory manual**. Cold Spring Harbor Laboratory; 1989.

Sander M, Sussel L, Connors J, et al. Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas. **Development** 2000; 127:5533-40.

Sanguinetti CL, Dias Neto E, Simpson AJG. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels **Biotechniques** 1994; 17:914-21.

Schott JJ, Benson DW, Basson CT, et al. Congenital heart disease caused by mutations in the transcription factor NKX2-5. **Science** 1998; 281:108-11.

Schwartz RJ, Olson EN. Building the heart piece by piece: modularity of cis-elements regulating Nkx2-5 transcription. **Development** 1999; 126:4187-92.

Tanaka M, Wechsler SB, Lee IW, Yamasaki N, Lawitts JA, Izumo S. Complex modular cis-acting elements regulate expression of the cardiac specifying homeobox gene Csx/Nkx2.5. **Development** 1999; 126:1439-50.

Teillet MA, Lapointe F, Le Douarin NM. The relationships between notochord and floor plate in vertebrate development revisited. **Proc Natl Acad Sci U S A** 1998; 95:11733-8.

Tenen DG, Hromas R, Licht JD, Zhang DE. Transcription factors, normal myeloid development, and leukemia. **Blood** 1997; 90:489-519.

The human genome. **Nature** 2001; 409:745-964.

9. Anexos

Este trabalho recebeu, em 2001, dois prêmios em diferentes congressos:

1. Primeiro Lugar no Concurso de Pôsteres da Revista Nature durante o “Brazilian International Genome Conference”, 26-29 de março em Angra dos Reis – RJ.
2. Prêmio Paineis 2001 da Sociedade Brasileira de Genética, durante o 47.º Congresso Nacional de Genética, 02-05 de outubro em Águas de Lindóia – SP.

Cópias dos prêmios seguem em anexo.



Brazilian International
Genome Conference



Covering Human, Mammalian, Plant and Microbial Genomics

26 – 29 March, 2001
Angra dos Reis,
Rio de Janeiro, Brazil.

2 year subscription to *Nature*

Awarded to

Poster Prize



Gaëlle Chopin Stukart

Signed

Philip Campbell

Philip Campbell
Editor, *Nature*

nature
www.nature.com



SBC Sociedade Brasileira de Genética

47º Congresso Nacional de GENÉTICA

Prêmio Painei 2001

Perkin Elmer – Life Sciences
Sociedade Brasileira de Genética

GAELE CHOPIN STUKART

GENÉTICA E EVOLUÇÃO HUMANA E GENÉTICA MÉDICA

Águas de Lindóia SP, 05 de outubro de 2001


Horácio Schneider
Presidente da Sociedade Brasileira de Genética - SBC


Angela V. Morgante
1ª Secretária da Sociedade Brasileira de Genética - SBC

