

**Universidade do Porto**



**Pesquisa de mutações no gene Six4  
em carcinoma ductal *in situ* da mama.**

**Mestrado em Oncologia Molecular**

**Elisabete Simões de Figueiredo**

**Porto, 2008**

## AGRADECIMENTOS

Agradeço a inspiração (a vossa força, coragem e sucesso), o carinho, o apoio incondicional e o incentivo

**aos meus pais e manita,**

(sem vocês não teria conseguido)

Agradeço a amizade, paciência *e algo mais que não sei dizer*

**aos meus pais**

**e aos meus amigos**

Eugénia, André, Inês,

Catarina, Paula, Misa,

Helena e Marta.

Agradeço o companheirismo laboratorial

à Ângela, à Catarina, à Inês, à Isabel, ao Luís e ao Nuno,

e a todos os *IPATIMUPianos*.

Agradeço a disponibilidade, a partilha de bibliografia e imagens, o apoio bio-informático e o apoio na sequenciação, respectivamente,

à Angélica e Nair, ao Filipe e João e à Mafalda.

Agradeço a oportunidade de descoberta de um campo de investigação extremamente interessante, que é a Oncologia Molecular, aos coordenadores e professores do Mestrado.

Agradeço colaboração do laboratório do Prof. Buckhaults, da Universidade de Carolina do Sul, e em particular a Randal L. Croshaw.

Agradeço as discussões, a partilha de conhecimento e sobretudo a orientação científica ao

Doutor Luís Teixeira da Costa.

Agradeço a oportunidade de integrar o IPATIMUP ao

Professor Sobrinho Simões

DISSERTAÇÃO DE CANDIDATURA AO GRAU DE MESTRE  
APRESENTADA À FACULDADE DE MEDICINA DA  
UNIVERSIDADE DO PORTO

Artigo 48º, § 3

– A Faculdade não responde pelas doutrinas expendidas na dissertação (Regulamento da Faculdade de Medicina do Porto – Decreto nº 19337, 29 de Janeiro de 1931)

## RESUMO

O carcinoma da mama é a neoplasia maligna mais comum em mulheres e uma das principais causas de morte por cancro (Ferlay *et al*, 2007).

Um estudo pioneiro (Sjöblom T *et al*, 2006) revelou que existe ainda um grande número de genes que não foram associados a neoplasias e que poderão fornecer novas indicações sobre a tumorigénese de carcinomas de tipo epitelial. A análise sistemática das alterações somáticas nos genomas de carcinoma da mama e colo-rectal levou à identificação de 189 genes, com uma grande probabilidade de terem sido sujeitos a selecção mutacional durante a tumorigénese, que foram denominados *Candidate Cancer Genes*, *CAN*. A identificação destes genes poderá levar a novas vias de investigação básica da biologia tumoral e conseqüentemente ter implicações ao nível de prognóstico, diagnóstico e de alvos terapêuticos.

As percentagens de mutações por substituição de uma única base encontradas nos carcinomas da mama e colo-rectal foram muito semelhantes, porém o espectro e o contexto nucleotídico entre os dois carcinomas foram bastante diferentes, o que poderá ser um reflexo das diferenças nos mecanismos de tumorigénese destes carcinomas. Em cada carcinoma da mama foram identificadas, em média, 52 mutações somáticas não-sinónimas, das quais 9 eram em genes *CAN*. Neste carcinoma foram detectadas mutações no *SIX4*, um gene de regulação de transcrição, que não foi anteriormente associado à neoplasia e que poderá estar implicado no processo de tumorigénese.

O *SIX4* é um membro da família *SIX*, pertencente ao grupo de genes homeobox divergentes ou de classe II e que apresenta uma estrutura modular, na qual cada exão corresponde a funções concretas. No *SIX4* humano, os exões parecem estar organizados em três grupos: o exão 1, que codifica o domínio *Six* e o homeodomínio de tipo *Six*, um segundo exão com função desconhecida e o exão3 que codifica um domínio de activação de transcrição (Ozaki *et al*, 1999).

Esta tese de dissertação pretende contribuir para o esclarecimento do eventual papel do *SIX4* na neoplasia da mama. Com esse propósito, propusemo-nos pesquisar alterações somáticas na região codificante do *SIX4* em 48 amostras clínicas de pacientes, diagnosticadas com carcinoma ductal da mama.

Nas amostras analisadas não foram detectadas alterações somáticas no SIX4, pelo que se pode especular que, como para outros genes homeobox (Goodman & Scambler, 2001), as mutações neste gene não são um evento frequente.

Foi detectada uma alteração não-sinónima no exão 3. Embora esta variante alélica já tenha sido descrita e a frequência alélica seja variável entre diferentes populações, a hipótese de esta variante alélica poder causar predisposição para o carcinoma da mama é plausível. Uma vez que o exão 3 codifica o domínio de activação de transcrição do SIX4 é importante tentar esclarecer as eventuais consequências funcionais da alteração e também das mutações somáticas anteriormente detectadas (Sjöblom T *et al*, 2006). Com esse intuito, propusemo-nos a analisar as consequências funcionais das alterações encontradas, nomeadamente a nível da dosagem de expressão de proteína e da capacidade de transactivação do SIX4. A estratégia delineada para clarificar a função do SIX4 na neoplasia da mama envolveu o desenvolvimento de um vector repórter do grau de activação do complexo de transcrição e de vectores onde é incluído o cDNA de SIX4 humano de referência (com o alelo ancestral) ou com as alterações detectadas, que serão usados para transfectar linhas celulares humanas.

## ABSTRACT

Breast cancer is the most frequent neoplasm in women and one of the leading cause of death from cancer (Ferlay *et al*, 2007).

A landmark study (Sjöblom T, *et al*, 2006) has shown there is still a large number of genes not previously associated to neoplasia that might give us important clues about epithelial tumorigenesis. A systematic analysis of the somatic mutations in breast and colorectal cancers has led to the identification of 189 genes very likely to have been subject to mutational selection during tumorigenesis, that have been termed Candidate Cancer Genes, or CAN genes. The identification of these genes might lead to new avenues of research in basic tumor biology, and have implications at the levels of diagnosis, prognosis and therapeutic targets.

The fraction of mutations that were single base substitutions was similar in breast and colorectal cancers, but the spectrum and nucleotide contexts of the substitutions differed widely between the two tumor types, perhaps reflecting differences in the tumorigenic mechanisms that give rise to them. On average, each breast carcinoma harbored non-synonymous mutations in 52 genes, 9 of which were CAN genes. Among the genes mutated in this tumor type was SIX4, a gene involved in transcription regulation that had not been previously associated to tumor formation.

SIX4 is a member of the SIX family, which belongs to the class II – also called “divergent” – type of homeobox genes and appears to have a modular structure, with different exons coding for different functional regions of the protein. In human SIX4, for instance, exon 1 codes for the *Six* domain and for a *Six*-type homeodomain, exon 2 codes for a region of unknown function and exon 3 codes for the transactivation domain.

The purpose of the work reported in this thesis was to help to determine if, indeed, SIX4 plays a role in mammary cancer development. To this end, 48 clinical samples from patients diagnosed with ductal carcinoma of the breast for mutations in the coding sequence of SIX4 have been screened. No somatic alterations were detected in SIX4, implying that, as for other homeobox genes (Goodman & Scambler, 2001), mutations in SIX4 are not a common event.

A non-synonymous alteration was indeed found in exon 3, in several cases, but comparison to the corresponding normal samples demonstrated it is a hereditary variant. Although this allelic variant had already been described and found to have different frequencies in various populations, it is still plausible that it might contribute to breast cancer predisposition, since, as noted, exon 3 codes for SIX4's transactivation domain. It is therefore important to try to determine the functional impact of this variant – and also of the somatic alterations previously detected (Sjöblom T, *et al*, 2006) – on protein expression levels as well as on transcription activation. We have devised a strategy for functional testing the different SIX4 variants, based on the development of a reporter vector for SIX4 transactivation and vectors for constitutive high-level expression of SIX4 in human mammalian cells.

## ÍNDICE

<b>Abreviaturas</b>	11
<b>Introdução</b>	14 a 51
Oncologia Molecular	15
Os genes <i>CAN</i>	18
Os genes Homeobox	21 a 29
Os genes Six	25
O Six4	28
A mama	30
Carcinoma da mama	32 a 47
Classificação morfológica	32
Dados epidemiológicos	34
Factores de risco	35
Padrões de expressão de genes	37
Estadiamento e Prognóstico	41
Mecanismos de tumorigénese	46
Genes Six e carcinoma da mama	48
<b>Metodologia</b>	52
<b>Resultados &amp; Discussão</b>	61
<b>Investigação em curso &amp; Perspectivas futuras</b>	84
<b>Bibliografia</b>	91

## ABREVIATURAS

**AJCC:** American Joint Committee on Cancer

**AMFR:** Autocrine motility factor receptor

**Antp:** Antennapedia

**APC:** Adenomatous polyposis coli

**AR:** Amphiregulin

**ARE:** Na,K-ATPase  $\alpha$ 1 subunit gene regulatory element

**ATM:** Ataxia-telangiectasia mutated

**ATP1A1:** Na,K-ATPase  $\alpha$ 1 subunit gene

**BER:** Base-excision repair

**BRCA1 e 2:** Breast cancer 1 e 2

**CaMP:** Cancer Mutation Prevalence

**CAN:** Candidate Cancer Genes

**cDNA:** Complementar deoxyribonucleic acid

**DCIS:** Ductal carcinoma *in situ*

**CIN:** Chromosome instability

**CDX:** Caudal-related homeobox

**CHEK2:** Checkpoint Kinase 2

**Dach:** Dachshund homolog

**DNA:** Deoxyribonucleic acid

**EGFR:** Epidermal Growth Factor Receptor

**EHD:** Extended homeodomain

**EP1:** EPinephrine

**EPHA3:** Ephrin receptor

**ER:** Estrogen receptor

**Erb:** Erythroblastic

**Eya:** eyes absent

**FBXW7:** F box and 7 WD (tryptophan-aspartic acid) domain protein

**HB-EGF:** Heparin Binding EGF

**HER2:** Human Epidermal Growth Factor Receptor 2

**HD:** Homeodomain

**HP:** Hexapeptide

**HOM:** Homeotic

**LOH:** Loss of heterozygosity

**MMR:** Mismatch repair

**NER:** Nucleotide-excision repair

**NRG1 e NRG2:** NeuReGulin 1 e 2

**Pax:** Paired Box

**PCR:** Polymerase chain reaction

**PTEN:** Phosphatase and tensin

**PTPKD:** Protein tyrosine phosphatase receptor

**RAS:** RAt Sarcoma

**RNA:** Ribonucleic acid

**STK11:** Serine/threonine kinase

**Six:** Sine oculis homeobox

**so:** Sine oculis

**TCF4 :** T cell factor 4

**TCF7L2:** Transcription factor 7-like 2

**TGF- $\alpha$** : Transforming Growth Factor  $\alpha$

**TGFBRII**: Transforming growth factor beta receptor II

**TNM**: Tumor, Node, Metastasis

**TP53**: Tumor protein p53

**UTDL**: Unidade Terminal Ducto-Lobular

**VEGF**: Vascular endothelial growth factor

# **INTRODUÇÃO**

## Oncologia Molecular

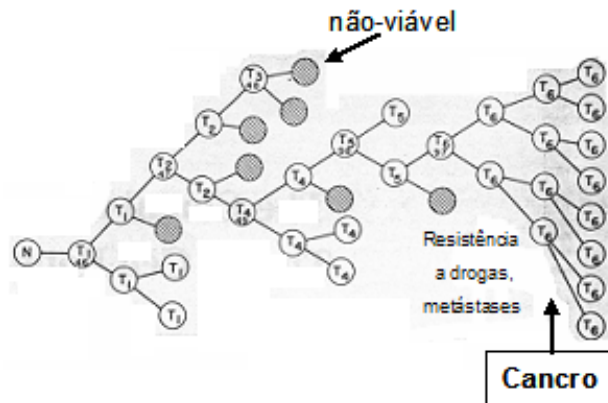
Uma das mais importantes descobertas na investigação da biologia do cancro e um dos dogmas centrais em Oncologia é que o cancro é uma doença genética (da Costa LF, 2001). Actualmente, acredita-se que, para atingir o estado maligno, as células dos tumores sólidos tiveram que acumular várias mutações em genes relevantes no processo neoplásico: os oncogenes, os genes supressores tumorais e os genes de estabilidade (Vogelstein B & Kinzler KW, 2004).

Os oncogenes são derivados dos proto-oncogenes, genes que promovem o crescimento normal. Nos tumores, os oncogenes com mutações encontram-se activos constitutivamente ou em condições em que o gene normal não estaria. A activação de oncogenes pode resultar de translocações cromossómicas, amplificações génicas ou mutações intragénicas subtis.

As mutações nos genes supressores tumorais levam à redução da actividade do produto génico. Para que as mutações nos genes supressores tumorais confirmem vantagem selectiva a uma célula é necessário que ambos os alelos (materno e paterno) sejam alterados, porém há autores (Santarosa M & Ashworth A, 2004) que sugerem que a inactivação de um único alelo é suficiente para que a célula adquira uma vantagem selectiva. A alteração de ambos os alelos pode ocorrer, por exemplo, através da perda de um braço ou do cromossoma inteiro, acoplada a uma mutação intragénica do outro alelo (Knudson AG, 2002). A inactivação dos genes supressores tumorais pode resultar de mutações *missense*, mutações *nonsense*, perdas ou inserções de vários tamanhos e silenciamento epigenético.

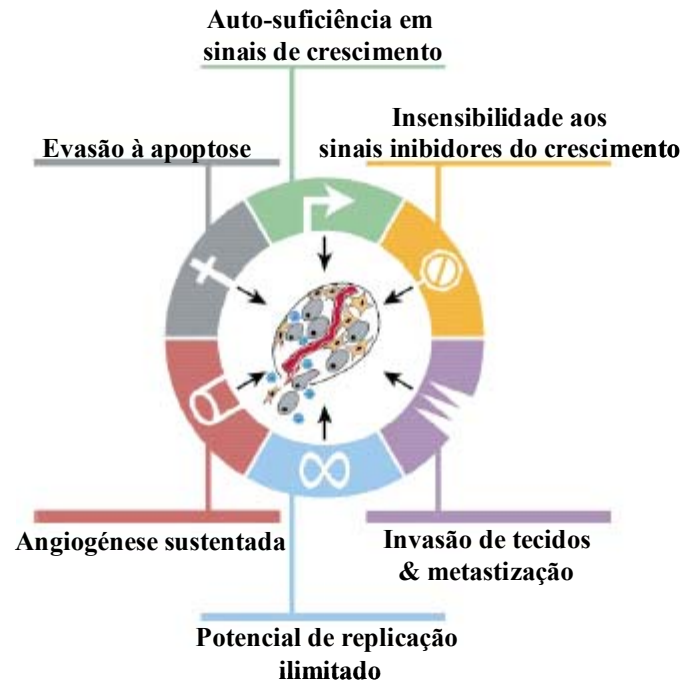
Os genes de estabilidade são responsáveis pela reparação de erros subtis que ocorrem durante a replicação normal de DNA ou induzidos pela exposição a agentes mutagénicos. Quando ocorrem deficiências num dos sistemas de reparação, a taxa de mutação aumenta consideravelmente. O aumento da taxa de mutação afecta potencialmente todos os genes, mas, apenas as mutações nos oncogenes ou genes supressores tumorais poderão conferir uma vantagem selectiva de crescimento às células. O processo neoplásico promovido por mutações nos oncogenes e genes supressores tumorais poderá ser provocado pela activação de genes que controlam o ciclo celular, pela inibição de processos apoptóticos normais ou pelo aumento da angiogénese (Vogelstein B & Kinzler KW, 2004).

Existem vários modelos que procuram explicar o mecanismo de tumorigênese (Nowell PC, 1976; Hanahan & Weinberg, 2000). De acordo com o modelo de evolução clonal (Nowell PC, 1976), um tumor inicia-se quando, através de uma alteração genética específica, uma célula adquire uma vantagem selectiva. A progressão tumoral é assegurada por sucessivas mutações e selecção clonal de variantes tumorais mais “anormais” e agressivas geradas continuamente ao longo do tempo (**Figura 1**).



**Figura 1:** Esquema que representa o modelo de evolução clonal de tumorigênese.  
(adaptado de Nowell PC, 1976)

Há autores (Hanahan & Weinberg, 2000) que sugerem que o crescimento maligno é consequência de seis novas capacidades adquiridas: auto-suficiência em sinais de crescimento, insensibilidade a sinais inibidores do crescimento, fuga à apoptose, potencial de replicação ilimitado, angiogênese sustentada e invasão de tecidos e metastização (**Figura 2**). Durante o desenvolvimento tumoral, cada uma destas alterações fisiológicas representaria uma barreira ultrapassada nos mecanismos de defesa das células e tecidos contra o cancro.



**Figura 2:** Capacidades adquiridas do cancro  
 (adaptado de Hanahan & Weinberg, 2000)

## Os genes *CAN*

A identificação de alterações genéticas é um dos indicadores mais fiáveis da potencial relevância de um gene na neoplasia humana (Vogelstein B & Kinzler KW, 2004). A determinação da sequência do genoma humano e o desenvolvimento de recursos como a sequenciação e a análise bioinformática permitem identificar, de forma exaustiva e imparcial, genes com potencial papel na tumorigénese do cancro humano (Sjöblom T, *et al*, 2006).

Os carcinomas da mama e colo-rectal têm em comum uma importância clínica considerável e a origem epitelial. A análise sistemática das alterações somáticas nos genomas destes carcinomas (mama e colo-rectal) levou à identificação de 236 genes com alterações somáticas. As mutações somáticas nos tumores humanos podem ser provenientes da selecção de alterações funcionalmente importantes ou da acumulação de alterações irrelevantes que ocorreram ao longo das sucessivas divisões celulares no tumor ou nas células progenitoras estaminais. Dos 236 genes identificados com alterações somáticas, os que apresentam uma grande probabilidade de terem sido sujeitos a selecção mutacional durante a tumorigénese foram denominados *Candidate Cancer Genes – CAN*. Verificou-se que cada tumor individual tinha a sua própria assinatura de mutações de genes *CAN*, e que nenhum dos tumores tinha mais de seis genes *CAN* com mutações em comum com qualquer outro (Sjöblom T, *et al*, 2006). Foram detectadas mutações em genes anteriormente associados a carcinoma da mama ou colo-rectal, em genes em que a prevalência de mutações em cancros esporádicos era relativamente baixa, em genes que já tinham sido associados a outros tipos tumorais (que não os cancros da mama e colo-rectal), em genes que tinham sido associados a neoplasia com base em estudos funcionais e de expressão e ainda em genes identificados como alvo de translocações ou amplificações. Por exemplo, descobriram-se mutações em *TP53*, *APC*, *KRAS*, *SMAD4*, *FBXW7 (CDC4)*, *EPHA3*, *TCF7L2 (TCF4)*, *BRCA1* e *TGFBR2*, *CDH10* e *CDH20* (caderinas), *PTPKD* (protein tyrosine phosphatase receptor), *AMFR* (autocrine motility factor receptor), entre outros. A re-identificação de todos os genes encontrados com mutações em mais de 10% de carcinomas da mama e colo-rectal constituiu uma validação crítica da abordagem usada (Sjöblom T, *et al*, 2006).

A tumorigénese do cólon é um dos processos melhor esclarecidos, nomeadamente no que diz respeito às lesões precursoras e sequências de eventos que levam ao estado maligno. O modelo proposto (Cho KR & Vogelstein B, 1992) para explicar a base genética da neoplasia colorectal salienta, entre outras características, que embora as alterações genéticas possam ocorrer numa sequência preferida, as propriedades biológicas do tumor são determinadas pela acumulação total das alterações e não pela sua ordem cronológica de aparecimento. Embora a percentagem de alterações por substituição de uma base, nos carcinomas da mama e colo-rectal, fosse semelhante, o espectro e o contexto nucleotídico eram bastante diferentes especulando-se que tal seja um reflexo de diferenças nos mecanismos de tumorigénese (Sjöblom T, *et al*, 2006).

Este estudo pioneiro (Sjöblom T, *et al*, 2006) revelou que existe ainda um grande número de genes, que não tinham sido associados a neoplasias, e que poderão fornecer novas indicações sobre a tumorigénese de carcinomas de tipo epitelial (**Tabela1**). A análise dos genes *CAN* poderá permitir abrir novas vias de investigação básica da biologia tumoral e conseqüentemente ter implicações ao nível de prognóstico, diagnóstico e de identificação de alvos terapêuticos.

Um dos genes *CAN* em que foram detectadas mutações e que não tinha sido anteriormente associado a estas neoplasias foi o SIX4. O SIX4 é um factor de transcrição pertencente ao grupo de genes homeobox de classe II ou divergente.

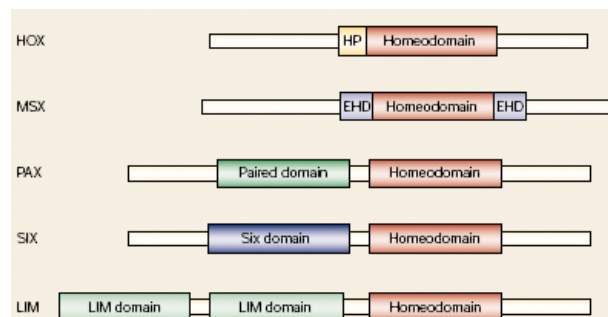
Breast cancers				Colorectal cancers					
Cellular adhesion and motility (examples: cytoskeletal protein binding GO:0008092, cell adhesion GO:0007155, metalloproteinase activity GO:0008237)									
<i>FLNB</i>	3.4	<i>TMPPRSS6</i>	2.0	<i>RAPH1</i>	1.4	<i>PKHD1</i>	3.5	<i>CNTN4</i>	1.6
<i>MNH1</i>	2.7	<i>COL11A1</i>	1.8	<i>PCDH815</i>	1.4	<i>ADAMTSL3</i>	3.3	<i>CNL1</i>	1.3
<i>SPTAN1</i>	2.6	<i>DNAH9</i>	1.7	<i>CHYA1</i>	1.4	<i>OBSCN</i>	3.0	<i>HAPLN1</i>	1.2
<i>DBN1</i>	2.5	<i>OBSCN</i>	1.7	<i>MACF1</i>	1.3	<i>ADAMTSL8</i>	2.7	<i>MGC33407</i>	1.2
<i>TECTA</i>	2.4	<i>COL7A1</i>	1.5	<i>SYNE2</i>	1.3	<i>MMP2</i>	2.3	<i>MAP2</i>	1.0
<i>ADAM12</i>	2.3	<i>MAGBE1</i>	1.5	<i>NRCAM</i>	1.1	<i>TTL3</i>	2.2		
<i>GSN</i>	2.2	<i>CDH10</i>	1.5	<i>COL19A1</i>	1.1	<i>EVL</i>	2.0		
<i>CDH20</i>	2.2	<i>SULF2</i>	1.5	<i>SEMASB</i>	1.1	<i>ADAM29</i>	2.0		
<i>BGN</i>	2.1	<i>CNTN6</i>	1.4	<i>ITGA9</i>	1.1	<i>CSMD3</i>	1.9		
<i>KAM5</i>	2.1	<i>THBS3</i>	1.4			<i>ADAMTSL5</i>	1.8		
Signal transduction (examples: intracellular signaling cascade GO:0007242, receptor activity GO:0004872, GTPase regulator GO:0030695)									
<i>VEPNI</i>	2.1	<i>PRC</i>	1.5	<i>PRPF4B</i>	1.3	<i>APC</i>	>10	<i>PTPRD</i>	2.2
<i>SBNO1</i>	2.1	<i>GAB1</i>	1.5	<i>CENTG1</i>	1.3	<i>KRAS</i>	>10	<i>MCP</i>	2.1
<i>DNASE1L3</i>	1.9	<i>ARHGEF4</i>	1.4	<i>MAP3K6</i>	1.3	<i>EPHA3</i>	4.2	<i>NF1</i>	1.9
<i>RAP1GA1</i>	1.8	<i>NALP8</i>	1.4	<i>APC2</i>	1.3	<i>GUCY1A2</i>	3.5	<i>PTPRU</i>	1.4
<i>BGFL6</i>	1.8	<i>RGL1</i>	1.4	<i>STARD8</i>	1.2	<i>EPHB6</i>	3.5	<i>CD109</i>	1.3
<i>AMFR</i>	1.7	<i>PPM1E</i>	1.4	<i>PTPN14</i>	1.1	<i>TGFB2</i>	2.9	<i>PHIP</i>	1.2
<i>CENTB1</i>	1.7	<i>PADRE1</i>	1.4	<i>IRTA2</i>	1.1	<i>GNAS</i>	2.6		
<i>GPNMB</i>	1.7	<i>CANM4</i>	1.3	<i>RASGRF2</i>	1.1	<i>RET</i>	2.3		
<i>INHBE</i>	1.7	<i>ALS2CL</i>	1.3	<i>MTMR3</i>	1.1	<i>P2RY14</i>	2.2		
<i>FLJ30458</i>	1.6	<i>RASAL2</i>	1.3			<i>LGR6</i>	2.2		
Transcriptional regulation (examples: regulation of transcription GO:0045449, zinc finger C2H2-subtype IPR007066)									
<i>TP53</i>	>10	<i>CHD5</i>	1.8	<i>ZFP64</i>	1.4	<i>TP53</i>	>10	<i>ZNF442</i>	1.9
<i>ELN2479</i>	3.4	<i>CIC</i>	1.7	<i>ZNF569</i>	1.4	<i>SMAD4</i>	4.6	<i>SMAD3</i>	1.9
<i>SIN4</i>	2.5	<i>KEAP1</i>	1.6	<i>ENMT1</i>	1.3	<i>MLL3</i>	3.7	<i>EY44</i>	1.5
<i>KIAK0934</i>	2.5	<i>HOXA3</i>	1.6	<i>ZFYVE26</i>	1.2	<i>TBX22</i>	3.3	<i>PKNOX1</i>	1.4
<i>LRRFP1</i>	2.4	<i>TCF1</i>	1.6	<i>BCLL1A</i>	1.1	<i>SMAD2</i>	3.1	<i>MKRN3</i>	1.3
<i>GLI1</i>	2.3	<i>HDAC4</i>	1.6	<i>ZNF318</i>	1.1	<i>TCF7L2</i>	2.8		
<i>RFX2</i>	2.1	<i>MYOD1</i>	1.5			<i>HIST1H1B</i>	2.5		
<i>ZCQ3</i>	1.8	<i>NCOM6</i>	1.5			<i>RUNX1T1</i>	2.4		
Transport (examples: ion transporter activity GO:0015075, ligand-gated ion channel activity GO:0015276, carrier activity GO:0005386)									
<i>ATP8B1</i>	3.1	<i>ABCB8</i>	1.7	<i>ABCB10</i>	1.4	<i>ABCA1</i>	2.8	<i>C6orf29</i>	1.1
<i>CUBN</i>	2.5	<i>KPNAS</i>	1.7	<i>SCNN1B</i>	1.3	<i>SLC29A1</i>	1.9		
<i>GRIN2D</i>	2.4	<i>ABCA3</i>	1.7	<i>NUP133</i>	1.1	<i>SCN3B</i>	1.9		
<i>NDLBP</i>	2.2	<i>SLC9A2</i>	1.6			<i>P2RX7</i>	1.3		
<i>NUP214</i>	1.8	<i>SLC6A3</i>	1.5			<i>KCNQ5</i>	1.2		
Cellular metabolism (examples: aromatic compound metabolism GO:0006725, generation of precursor metabolites GO:0016445, biosynthesis GO:0009058)									
<i>ACADM</i>	2.0	<i>NCB50R</i>	1.7	<i>PHACS</i>	1.4	<i>UQCRC2</i>	1.9		
<i>PRPS1</i>	1.8	<i>ASL</i>	1.6	<i>XDH</i>	1.3	<i>ACSL5</i>	1.6		
<i>CYP11A1</i>	1.7	<i>GALNT5</i>	1.4			<i>GALNS</i>	1.2		
Intracellular trafficking (examples: endoplasmic reticulum targeting sequence IPR000866, membrane fusion GO:0006944)									
<i>OTOF</i>	2.2	<i>PLEKHAB</i>	1.8	<i>KTN1</i>	1.5	<i>SYNE1</i>	2.3	<i>PRAD1</i>	1.9
<i>LRBA</i>	2.1	<i>LOC283849</i>	1.7	<i>GGA1</i>	1.4	<i>SEC8L1</i>	2.2	<i>LRP2</i>	1.2
<i>AEGP</i>	1.8	<i>SORL1</i>	1.7			<i>SDCAG84</i>	2.2		
RNA metabolism (examples: RNA processing GO:0008353, RNA splice site selection GO:0006376)									
<i>C14orf155</i>	3.3	<i>RNU31P2</i>	1.7	<i>KIAA0427</i>	1.5	<i>SFRS6</i>	1.3		
<i>SP110</i>	1.8	<i>C22orf9</i>	1.5	<i>DDX10</i>	1.3				
Other (examples: response to DNA damage stimulus GO:0006974, protein ubiquitination GO:0016567)									
<i>FLJ0869</i>	2.1	<i>SERPINB1</i>	1.4			<i>FBXW7</i>	5.1	<i>K6IRS3</i>	1.2
<i>BRCA1</i>	2.0					<i>UHRF2</i>	1.5	<i>CD248</i>	1.2
<i>MRE11A</i>	1.6					<i>LMO7</i>	1.3	<i>ERCC6</i>	1.0
Unknown									
<i>KIAA1632</i>	2.4	<i>KIAA0999</i>	1.3			<i>C10orf237</i>	2.7	<i>KIAA1409</i>	1.6
<i>MGC24047</i>	2.1					<i>LOC157697</i>	2.0	<i>C15orf2</i>	1.0

TABELA 1: Genes *CAN* e sua classificação funcional

## Os genes homeobox

Os genes homeobox codificam uma família de factores de transcrição que desempenham um papel fundamental na embriogénese (McGinnis & Krumlauf, 1992).

A família de genes homeobox é caracterizada pelo motivo homeobox (McGinnis *et al*, 1984) que foi inicialmente identificado em *Drosophila* como sendo uma sequência partilhada pelos genes homeóticos – do complexo HOM-C. O seu papel foi estabelecido quando uma mutação no gene Antennapedia (Antp) da *Drosophila* levou à formação, em vez de antenas, de um par de pernas na cabeça (mutação homeótica) (Struhl, 1981). Actualmente sabe-se que o motivo homeobox está presente em muitos genes (**Figura 3**) e virtualmente em todas as espécies eucarióticas. Os genes HOM-C continuam a ser considerados os genes homeobox protótipo e os seus homólogos em humanos – o complexo HOX – estão entre os genes homeobox mais estudados em vertebrados. Apesar disso, o grupo HOX representa apenas um pequeno grupo de todos os genes homeobox conhecidos – e já foram descobertos mais de 200 nos vertebrados (Tupler R *et al*, 2001).



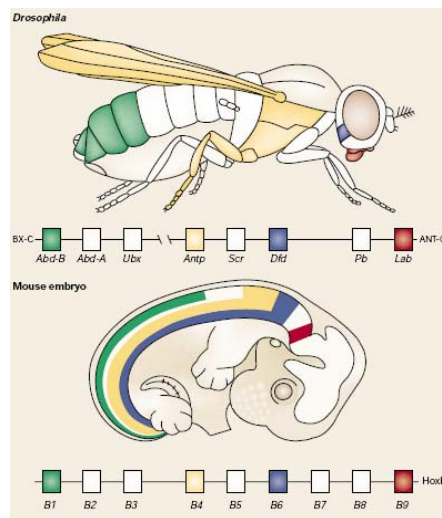
**Figura 3:** As diferentes classes de genes homeobox. HP: hexapeptídeo, EHD: extended homeodomain.

(adaptado de Abate-Shen C, 2002)

De acordo com a homologia dos respectivos homeodomínios, os genes homeobox estão divididos em dois grandes grupos, os de classe I e os de classe II ou divergentes (revisto em Chen & Sukumar, 2003).

Os genes homeobox de classe I partilham uma elevada homologia entre os homeodomínios com o primeiro gene – Antp – descoberto em *Drosophila*. Nos

humanos, os genes homeobox de classe I são designados HOX e estão organizados em quatro grupos, localizados nos cromossomas 2, 7, 12, e 17. Cada grupo contém no mínimo nove genes HOX. Os genes HOX são amplamente expressos durante a embriogénese e geralmente a sua expressão restringe-se a células não diferenciadas ou em proliferação (revisto em Abate-Shen C, 2002). A sua ordem de expressão ao longo do eixo antero-posterior no embrião (**Figura 4**) é colinear com a organização dos genes nos cromossomas. Cada órgão específico expressa um grupo de genes HOX e os seus níveis de expressão variam ao longo das diferentes fases de desenvolvimento do órgão.

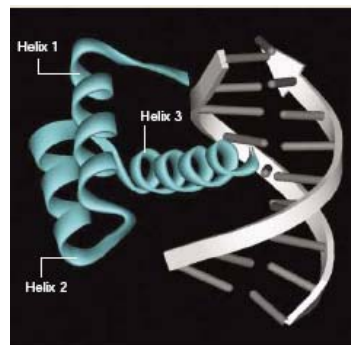


**Figura 4:** A expressão colinear dos genes HOM-C e HOX. Os genes que estão localizados na extremidade 5' do cromossoma são expressos na zona posterior do embrião, enquanto os que se localizam na extremidade 3' são expressos na zona anterior. (Abate-Shen C, 2002)

Os genes homeobox divergentes ou de classe II apresentam uma homologia com o gene *Antp* inferior a 50%. Os genes homeobox divergentes são agrupados em várias subfamílias de acordo com a presença de sequências adicionais, como: SIX, MSX, PAX, entre outras. Além do homeodomínio, os motivos adicionais conservados contribuem para as propriedades funcionais distintas dos membros individuais das famílias homeobox. Ao contrário dos genes HOX, a maioria dos genes homeobox encontra-se disperso ao longo do genoma (revisto em Chen & Sukumar, 2003).

O homeodomínio é constituído por 60 resíduos de aminoácidos, que formam uma estrutura com três hélices (Gehring WJ *et al*, 1994) (**Figura 5**). Apesar de se ter demonstrado que muitas homeoproteínas actuam como factores de transcrição,

reconhecendo sequências de DNA como TGAT, a TAAT, A TTAT e TTAC, o número de genes alvo *in vivo* identificados é reduzido. Uma das dificuldades na identificação dos genes alvo é a promiscuidade da ligação das homeoproteínas *in vitro*. Embora seja aceite que a promiscuidade da ligação *in vitro* reflecte a necessidade de co-factores *in vivo*, também foi sugerido que a baixa especificidade pode implicar que as homeoproteínas actuam simultaneamente como reguladores de múltiplos genes alvo (revisto em Abate-Shen C, 2002).



**Figura 5:** Estrutura tri-dimensional do homeodomínio.

(Abate-Shen C, 2002)

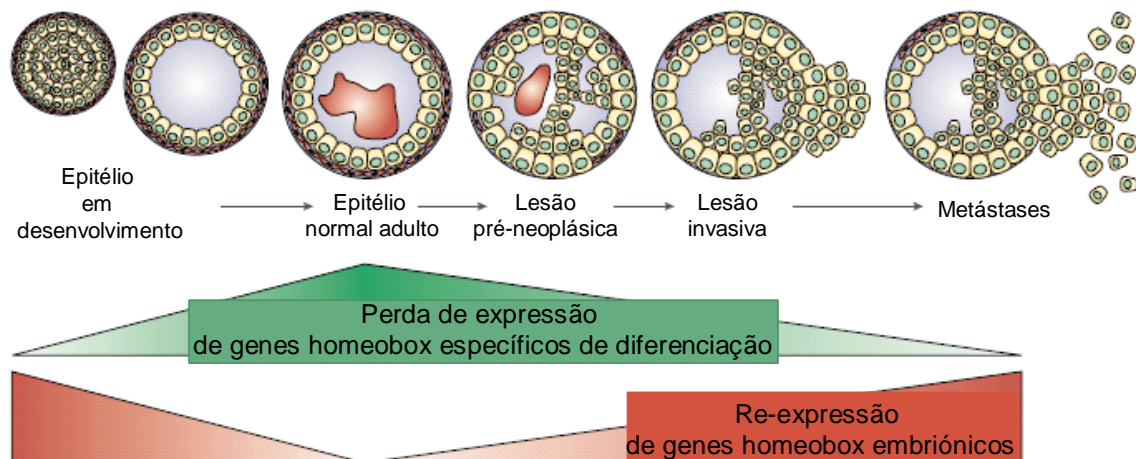
Em condições normais, a actividade das homeoproteínas mantém um equilíbrio entre a proliferação e diferenciação, mas quando a sua expressão é perturbada poderá ser adquirido um potencial para promover o processo de tumorigénese, pelo que foi proposto que os genes homeobox sejam definidos como “moduladores tumorais” (Abate-Shen C, 2002).

A expressão dos genes homeobox encontra-se desregulada em vários tumores sólidos e nas linhas celulares deles derivadas. Consoante a alteração da expressão, podem considerar-se três categorias:

- 1) genes que podem ser re-expressos em células neoplásicas provenientes de células embrionárias nas quais o gene era expresso durante o desenvolvimento,
- 2) genes que podem ser expressos em células neoplásicas provenientes de células embrionárias nas quais o gene não era expresso,
- 3) genes que podem ser sub-expressos em células neoplásicas provenientes de tecidos nas quais o gene é normalmente expresso no estado diferenciado.

Generalizando, os genes homeobox que são “ganhos” ou sobre-expressos em carcinomas apenas são expressos em células indiferenciadas ou em proliferação. Do

mesmo modo, os genes expressos em condições normais nos tecidos diferenciados geralmente são “perdidos” ou sub-expressos em carcinomas. Em termos gerais, poderá considerar-se que o estado de diferenciação celular é estabelecido pela expressão contínua de determinados genes homeobox e pela ausência de outros e que o fenótipo tumoral é uma consequência da sua expressão aberrante (**Figura 6**).



**Figura 6:** A relação entre a expressão dos genes homeobox, o desenvolvimento epitelial e a tumorigênese.

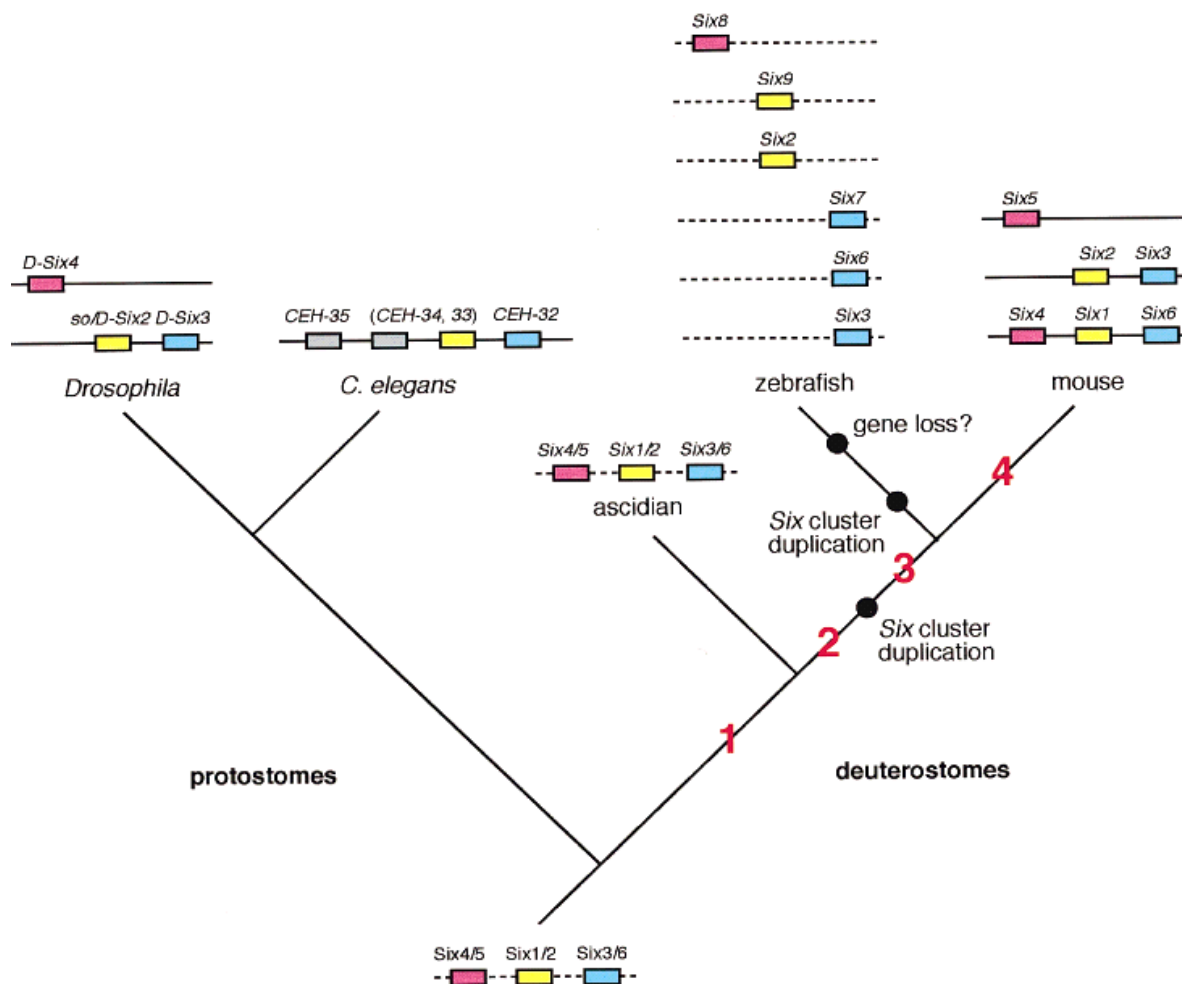
(Abate-Shen C, 2002)

As diferentes classes de genes homeobox assumem uma relevância particular de acordo com a sua expressão espacial e temporal (revisto em Chen H & Sukumar S, 2003). Uma vez que a função desempenhada pelas homeoproteínas também depende da expressão espacial foi sugerido que o seu potencial oncogénico seja consequência da sua expressão normal num contexto celular desadequado.

## Os genes *Six*

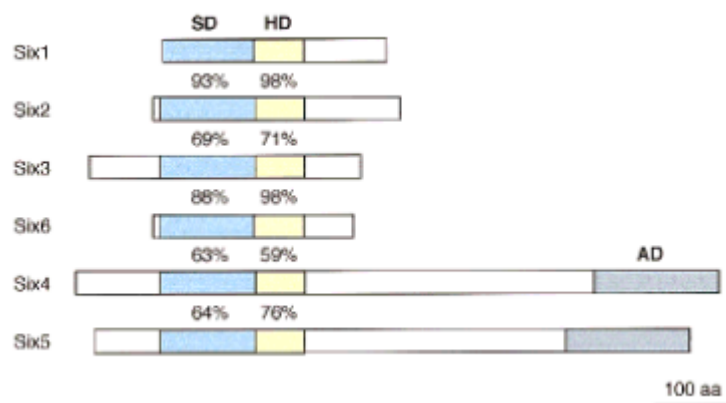
Os genes *Six* constituem uma sub-família de genes homeobox caracterizada pelo domínio *Six* adjacente ao homeodomínio. Os genes *Six* foram inicialmente identificados como homólogos do gene *sine oculis* (*so*), que é essencial no desenvolvimento do sistema visual de *Drosophila* (Cheyette *et al*, 1994) e que é regulado por uma via genética conservada entre invertebrados e vertebrados (Seo *et al*, 1999).

A caracterização dos genes *Six* em várias espécies mostra que esta sub-família se encontra conservada em vários metazoários e as análises filogenéticas sugerem que o genoma do ancestral comum dos insectos e vertebrados continha 3 genes *Six* diferentes a partir dos quais se originaram os diferentes subgrupos (Seo *et al*, 1999) (**Figura 7**).



**Figura 7:** Esquema que representa o modelo de evolução proposto para os genes *Six*.  
(Kawakami *et al*, 2000)

As proteínas Six apresentam regiões amino e carboxi terminais altamente variáveis e dois domínios adjacentes conservados, o Six (SD, 110-115 aa) e o homeodomínio de tipo Six (HD, 60 aa). O homeodomínio da família Six é caracterizado pela ausência de dois resíduos de aminoácidos que são altamente conservados na maioria dos homeodomínios – um resíduo de arginina na posição 5 e um de glutamina na posição 12, que se localizam nas regiões amino-terminal e na hélice 1, respectivamente. A substituição destes dois resíduos de aminoácidos pode levar à modificação da estrutura e afectar a especificidade da ligação ao DNA (Laughon, 1991; Serikaku *et al*, 1994). Apesar desta característica saliente, existe uma considerável divergência entre as sequências dos HDs, e maior ainda entre os SDs, dos diferentes membros da família Six. Com base na homologia entre os domínios SD e HD foi sugerida uma classificação em três subgrupos principais, de acordo com a semelhança com o Six2, o Six3 ou o Six4. Cada membro individual é incluído num desses subgrupos principais e possui um tetrapéptido semelhante junto à extremidade amino do HD. Ao contrário do subgrupo do Six3 que apresenta um tetrapéptido (QKTH) bastante distinto, existe uma grande semelhança entre os dos subgrupos do Six2 (ETSY) e do Six4 (ETVY) (Seo *et al*, 1999). Nos mamíferos foram identificados 2 membros para cada um dos 3 subgrupos Six sugeridos e foi proposto que se designassem Six1/2, Six3/6 e Six4/5 (**Figura 8**).



**Figura 8:** Estrutura da família de proteínas Six e homologia entre os SD e HD.

O domínio SD actua, quer na ligação a sequências específicas de DNA, quer em interacções proteína-proteína (kawakami *et al*, 1996 b; Pignoni *et al*, 1997). É possível que as diferenças de especificidade de ligação ao DNA ou a capacidade de cooperar e recrutar parceiros distintos possam influenciar a importância relativa de cada gene Six

num processo de desenvolvimento normal. Alguns membros retêm funções conservadas como componentes da via de regulação Pax-Six-Eya-Dach, que está envolvida no controlo da proliferação e movimento celular durante a embriogénese. A co-expressão de determinadas combinações de Six e Eya induzem a activação sinérgica do promotor, que está associada à translocação nuclear de co-activadores de Eya. Enquanto na ausência da co-expressão de Six, os co-factores Eya se localizam no citoplasma, a co-expressão de Six2, Six4 ou Six5, pode induzir a translocação nuclear de Eya1, Eya2 e Eya3. As diferenças evidentes entre as sequências específicas de ligação ao DNA das proteínas Six3 e Six 6 e os outros membros da família Six e o facto do Six3 não induzir a translocação de proteínas Eya, levam a sugerir que a subfamília Six3/6 coopere com outros co-factores e que não partilhem os mesmos genes alvo.

No ratinho, um dos mamíferos onde os *Six* se encontram melhor caracterizados, os genes Six1, Six2 e Six5 apresentam padrões de expressão vastos enquanto o Six3 e o Six6 são expressos exclusivamente no prosencéfalo e nos olhos em desenvolvimento. Embora tenha sido observada a expressão de mRNA de Six2, Six4 e Six5 na retina, a sua função no processo de desenvolvimento do olho é desconhecida, e no ratinho apenas o Six3 é expresso durante os estádios mais precoces desse processo. Os padrões de expressão do Six3 e do Six4, durante a embriogénese, sugerem um envolvimento na especificação ou na diferenciação de partes da cabeça, incluindo o cérebro, o sistema neuro-endocrinológico e vários órgãos sensoriais (Kawakami *et al*, 2000).

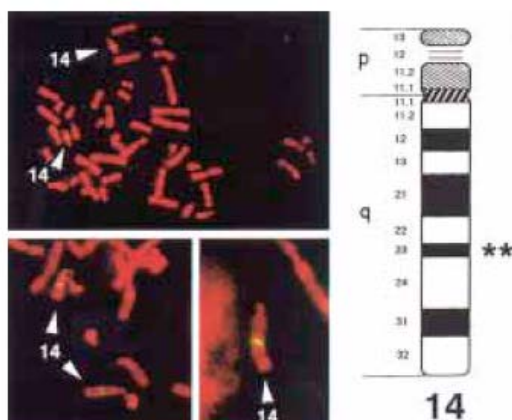
Os dados filogenéticos e os padrões de expressão cefálica dos membros individuais Six sugerem que esta família, originalmente, tinha como função a especificação das estruturas da cabeça. Os genes Six parecem ter em comum, no desenvolvimento de vários órgãos, a capacidade de estimular a proliferação das células com linhagens específicas antes da sua diferenciação final e a regulação do movimento das células durante a morfogénese. Considerando a regulação temporal e espacial da expressão dos genes Six, assim como a sua função como factores de transcrição, a identificação dos seus genes alvo é essencial para compreender os papéis desempenhados por cada um dos membros da família (Kawakami *et al*, 2000).

## O SIX4

Originalmente a proteína Six4 foi identificada como um factor de ligação a ARE (Na,K-ATPase  $\alpha$ 1 subunit gene regulatory element), um elemento de regulação da subunidade alfa de Na,K-ATPase, tendo sido denominada AREC3 (Kawakami K *et al*, 1996 a). A isoforma  $\alpha$ 1 (ATP1A1) é uma das subunidades catalíticas da enzima Na,K-ATPase, que é responsável pela manutenção do gradiente electroquímico dos iões  $\text{Na}^+$  e  $\text{K}^+$  da membrana celular das células animais, sendo extremamente importante na homeostase celular (Sweadner KJ, 1989). Esta enzima é composta por duas subunidades distintas, a alfa ( $\alpha$ ) e a beta ( $\beta$ ), cuja expressão é regulada diferencialmente nos vários tecidos e estádios de desenvolvimento. Ao contrário do que se passa com o Six3, as regiões que constituem os SD e HD do Six2 e Six 5 também se ligam ao ARE (Kawakami K *et al*, 2000).

As análises dos domínios mostraram que o Six4 apresentam duas substituições características no seu HD, um resíduo de lisina (Lys, posição 17) e um de arginina (Arg24), e que quer o SD quer o HD são responsáveis pela actividade de ligação específica ao DNA. Sugerem ainda que o Six4 actue como um factor de transcrição, já que a região carboxi-terminal contém um domínio de transactivação (Kawakami *et al*, 1996 a).

O gene SIX4 humano localiza-se no braço longo do cromossoma 14- 14q23 (Ozaki *et al*, 1999) (**Figura 9**), onde também se localizam, em *loci* muito próximos, os genes Six1 e Six6.



**Figura 9:** Localização cromossómica do gene SIX4 humano. O mapeamento do SIX4 em 14q23 é apresentado em imagens típicas de FISH.

(Ozaki *et al*, 1999)

O SIX4 tem uma estrutura modular na qual os diferentes exões correspondem a funções distintas: o exão 1 codifica o domínio *Six* e o homeodomínio, o exão 2 codifica uma região com função desconhecida e o exão 3 codifica um domínio de activação de transcrição. O exão 1, que especifica a actividade de ligação ao DNA, mantém-se ao longo da evolução da família de genes *Six* nos mamíferos. A região codificada pelo exão 3 é uma característica proeminente dos genes *Six4/5* em comparação com os outros membros, sugerindo uma função distinta para esta subfamília (Ozaki *et al*, 1999). O padrão espacial de expressão do *Six4* foi conservado ao longo da evolução dos vertebrados. A expressão específica em certas células neuronais durante a embriogénese, e em músculos esqueléticos no adulto, levam a especular que o SIX4 esteja envolvido na neurogénese, no desenvolvimento de certos órgãos e na miogénese (Ozaki *et al*, 2001). A activação do promotor do gene miogenina, por ligação directa ao MEF3 ou em sinergia com um co-factor específico, o Eya, levam a reforçar a ideia que o SIX4 é um dos genes envolvidos no controlo da miogénese (Ohto *et al*, 1999).

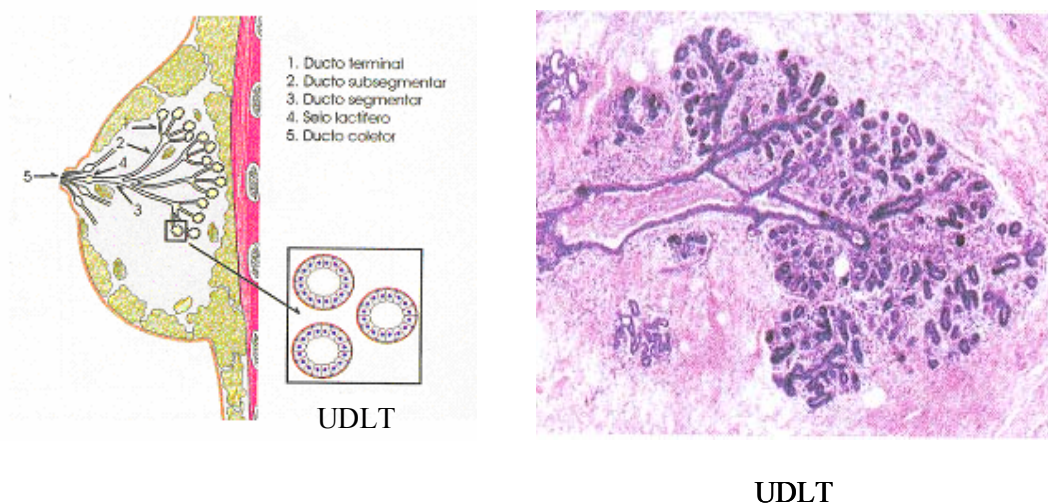
Foi proposto que um dos papéis do *so* pode ser a regulação dos genes que controlam os movimentos morfogenéticos no lóbulo óptico e noutros locais onde é expresso (Cheyette *et al*, 1994). Por analogia, o papel do *Six4* poderá ser o desempenho de funções morfogenéticas nas regiões em que é expresso.

A maturação da mama ocorre sobretudo após o nascimento e por isso a glândula mamária é um modelo ideal para o estudo de genes homeobox na tumorigénese. As diferentes fases de desenvolvimento da mama podem caracterizar-se como uma série de transições orquestradas, onde são tomadas decisões críticas relativamente: a diferenciação celular, à formação de estrutura e à função celular (Lewis, 2000). A estrutura e função da mama normal requerem interações complexas entre as células do estroma, as luminais e as mioepiteliais. Durante a tumorigénese, as mesmas funções que permitem a formação normal de novos ductos e alvéolos durante a puberdade e gravidez – eliminação da membrana basal, aumento da proliferação, angiogénese e invasão do estroma – poderão ser adoptadas pelas células epiteliais com anomalias (Wiseman BS & Werb Z, 2002).

## A mama

A mama é considerada uma glândula sudorípara modificada. A formação da mama é iniciada na vida intra-uterina mas, ao contrário de outros órgãos, o seu desenvolvimento ocorre sobretudo após a puberdade. Na puberdade feminina, com o início da secreção cíclica de estrogénio e progesterona, observa-se um alongamento ductal rápido e invasivo. Durante a gravidez, as alterações hormonais levam a uma transição drástica de uma morfologia predominantemente ductal para lóbulo-alveolar. A plenitude da maturação é atingida durante a gestação, ocorrendo involução após a paragem da lactação. O ciclo de desenvolvimento dá-se por concluído, aquando da menopausa, com a hipotrofia da mama (Lewis, 2000).

Na mulher, a glândula mamária (**Figura 10**) é constituída por 6 a 10 sistemas de ductos ramificados a partir do mamilo, que se estendem radialmente entre o estroma fibro-adiposo. No homem, a mama é rudimentar e inactiva, e o epitélio mamário encontra-se limitado aos grandes ductos junto ao mamilo. Na mulher adulta, o sistema de ductos ramificados da mama pode dividir-se em dois grupos, a Unidade Ducto-Lobular Terminal (UDLT) e os grandes ductos. A ramificação sucessiva dos ductos origina a UDLT e os lóbulos (ou alvéolos).

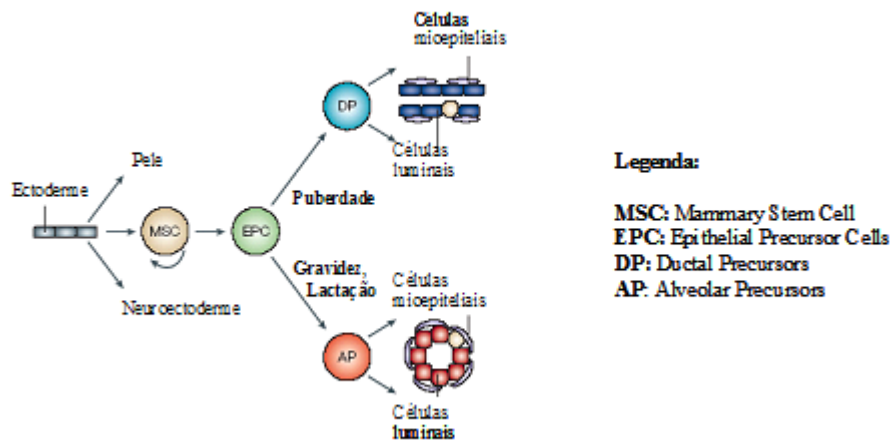


**Figura 10:** Microanatomia da mama; UDLT: Unidade Ducto-Lobular Terminal.

(adaptado de Schmitt F & Gobbi H, 2006)

O mamilo, também conhecido como papila mamária, contém, além dos grandes ductos, tecido conjuntivo denso, fibras musculares e numerosas glândulas sebáceas. A auréola é um anel de pele centrado no mamilo que contém glândulas sebáceas modificadas que se abrem na superfície em pequenas elevações, conhecidas como tubérculos de Montgomery.

Na mama normal, os ductos e lóbulos são revestidos por dois tipos de células epiteliais, as mioepiteliais e as luminais. As células mioepiteliais estão assentes na membrana basal e, além de auxiliarem na ejeção do leite durante a lactação, desempenham um papel importante na manutenção da estrutura normal e função do lóbulo e membrana basal (Lakhani & O' Hare, 2001). As células luminais são a camada mais interna e as que revestem o ductos terminais e os lóbulos produzem leite. Foi proposto (Hennighausen & Robinson, 2005; Bocker *et al*, 2002) que as células ductais ou lobulares são originadas a partir de células mamárias estaminais especializadas localizadas nos ductos terminais (**Figura 11**). Os lóbulos estão rodeados pelo estroma intralobular, que é sensível a hormonas específicas e que contém linfócitos. A maioria do estroma da mama, interlobular, consiste em tecido conectivo denso fibroso misturado com tecido adiposo.



**Figura 11:** Linhagens de células no epitélio mamário

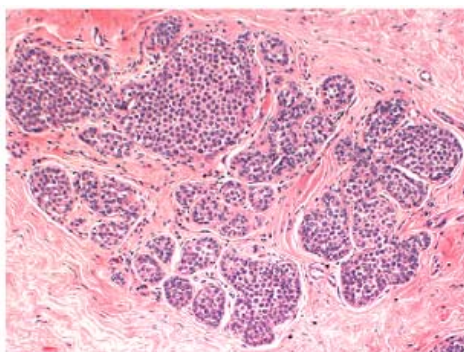
(Adaptado de Hennighausen & Robinson, 2005)

## Carcinoma da mama

### Classificação morfológica

Os carcinomas da mama são neoplasias epiteliais malignas, que se classificam, quanto à localização e extensão, em carcinomas *in situ* ou invasores. O carcinoma *in situ* consiste numa população de células neoplásicas confinadas aos ductos ou lóbulos e limitada pela membrana basal. O carcinoma invasor ou “infiltrante” distingue-se do carcinoma *in situ* pela invasão do estroma. No carcinoma invasor, as células neoplásicas podem invadir a vasculatura, dando origem a metástases nos nodos linfáticos regionais ou em locais distantes (Lester SC, 2005).

O carcinoma *in situ* foi, originalmente, dividido em lobular e em ductal: carcinoma lobular *in situ* (CLIS) e carcinoma ductal *in situ* (CDIS). Embora ainda se usem estes termos, a divisão não implica um local ou tipo celular de origem. A divisão em lobular e ductal baseia-se no padrão citológico e arquitetural das células proliferativas. O carcinoma lobular *in situ* (**Figura 12**) pode ser identificado pelos padrões morfológicos de infiltração característicos, com células únicas ou grupos de células fracamente ligadas (Lester SC, 2005).

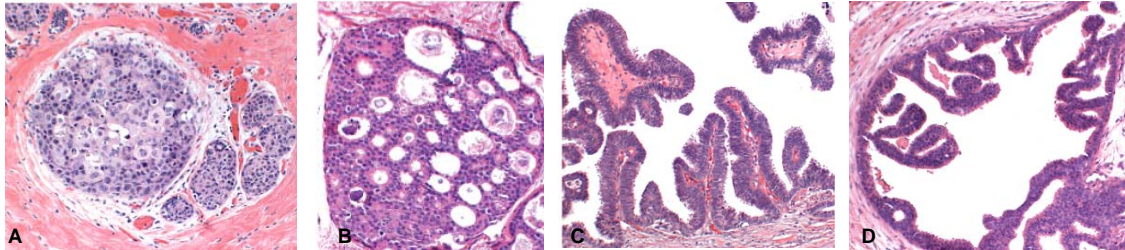


**Figura 12:** Carcinoma lobular *in situ* (CLIS).

(Lester SC, 2005)

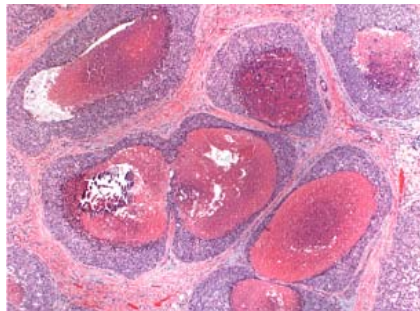
Historicamente, o CDIS foi dividido em cinco subtipos arquiteturais: sólido, cribriforme, papilar, micropapilar (**Figura 13, A-D**) e comedocarcinoma (**Figuras 14**). O termo comedocarcinoma refere-se a tumores de elevado grau nuclear associados a

extensa área de necrose. Geralmente o carcinoma ductal *in situ* envolve apenas um sistema ductal, no entanto as células em proliferação podem espalhar-se através dos ductos e lóbulos e originar lesões que envolvem um sector inteiro de uma mama.



**Figura 13:** Carcinoma ductal *in situ* (CDIS); Não-Comedo: A, sólido; B, cribriforme, C, papilar; D, micropapilar.

(Lester SC, 2005)



**Figura 14:** Carcinoma ductal *in situ* (CDIS); Comedocarcinoma.

(Lester SC, 2005)

## Dados epidemiológicos

O carcinoma da mama é a neoplasia maligna mais frequente em mulheres e uma das principais causas de morte por cancro (Ferlay *et al*, 2007).

A incidência do cancro da mama varia amplamente com a localização geográfica. Nos países industrializados, como os do Norte da América e Europa, registam-se as taxas de incidência de cancro da mama mais elevadas. O Japão surge como excepção, sendo um dos países industrializados que apresenta uma taxa de incidência mais moderada (Bray *et al*, 2004). Embora as taxas de incidência estejam a aumentar na maioria dos países, o prognóstico de cancro da mama, nos países ocidentais, é considerado bastante favorável (Parkin M *et al*, 2005).

Na maioria dos países europeus observou-se um declínio nas taxas de mortalidade por cancro da mama, sobretudo em mulheres jovens. Contudo, dado que o cancro é um processo cumulativo e a população europeia se encontra envelhecida, o número de mortes por cancro da mama continua a aumentar. Especula-se que a redução ou estabilização das taxas de mortalidade nos países com elevados recursos sócio-económicos, se deve ao efeito combinado de um diagnóstico precoce e de tratamentos com melhor qualidade (Ferlay *et al*, 2007). O diagnóstico precoce deve-se a uma maior sensibilização da sociedade para a doença e seus sintomas, e à aplicação de programas de detecção, como a mamografia, que permitem detectar precocemente os tumores, evitando a sua progressão. Os tratamentos com melhor qualidade reflectem os avanços no conhecimento da biologia do cancro, o que permite conceber terapias mais adequadas e eficazes.

## Factores de risco

O cancro da mama apresenta uma nítida etiologia multi-factorial, contudo os factores hormonais e genéticos são os de maior preponderância. O carcinoma da mama é uma neoplasia que raramente ocorre no homem, porém os factores de risco são semelhantes aos da mulher. Alguns dos factores de risco implicados no cancro da mama são:

- Factores endócrinos e reprodutivos: menarca precoce, menopausa tardia, primeira gestação tardia (após os 35 anos). Uma amamentação prolongada leva a uma redução do risco de cancro da mama.

- Factores genéticos ou História familiar: mutações no BRCA1 e BRCA2, no PTEN, no TP53, no STK11 e no ATM (Thull & Vogel, 2004). O risco de cancro da mama aumenta com o número de parentes de primeiro grau afectados. No homem, 3 a 8% dos casos estão associados ao Síndrome de Klinefelter.

- Carcinoma da mama contra-lateral ou do endométrio.

- Doença da mama benigna: biópsias da mama que apresentam hiperplasia atípica estão associadas a um risco elevado de cancro da mama.

- Idade: o tumor é muito raro antes dos 25 anos; a sua incidência aumenta a partir dos 30 anos e atinge um pico entre os 50-60 anos.

- Exposição a estrogénio exógeno.

- Exposição a radiação ionizante: o risco aumenta quando a exposição à radiação acontece numa idade mais precoce e em doses mais elevadas.

- Influência geográfica: a diferença observada na incidência dos diferentes países pode reflectir factores de risco modificáveis, como a dieta, o exercício físico, a obesidade e o consumo de álcool.

É provável que a predisposição para cancro da mama se deva ao efeito colectivo das variações num grande número de genes de susceptibilidade, com penetrância reduzida, (Antoniou *et al*, 2002; Peto, 2002), aliados à influência de factores ambientais, como por exemplo a exposição ao estrogénio.

As mutações em genes que podem contribuir para o processo neoplásico podem ocorrer em células somáticas, resultando em tumores esporádicos ou podem ocorrer na linha germinativa, resultando numa predisposição hereditária (Lester SC, 2005). A maior

percentagem de casos de cancro da mama são carcinomas esporádicos, contudo 5 a 10% de todos os casos resultam de uma predisposição hereditária.

Nos cancros da mama hereditários, salientam-se os genes BRCA1 ou BRCA2, que ilustram o paradigma clássico de genes supressores tumorais: embora a mutação de um dos alelos na linha germinativa seja suficiente para causar a predisposição para cancro é necessário que o segundo alelo seja perdido nas linhas somáticas (Venkitaraman, 2002). As mutações nos genes BRCA1 e BRCA2 podem ocorrer em qualquer sítio ao longo do gene, excepto no caso de certos grupos populacionais, como os judeus Ashkenazi, onde surgem três mutações específicas com elevada frequência: duas no gene BRCA1 (185 del AG e 5382 ins C) e uma no gene BRCA2 (6174 del T) (Tonin *et al*, 1996). Uma família que tenha pelo menos um homem afectado com cancro da mama, tem 60 a 70% de probabilidade de ter mutações no gene BRCA2. Nos tumores esporádicos, raramente se encontram mutações nos genes BRCA1 e BRCA2. No entanto, cerca de 50% destes tumores apresentam uma redução ou mesmo ausência da expressão de BRCA1.

Outros síndromas clínicos para os quais foi estabelecida uma predisposição genética para o cancro da mama, devido a mutações em genes específicos, são: o de Cowden, o de Li-Fraumeni, o de Peutz-Jeghens e a Ataxia-Telangiectasia, que são causados respectivamente, por mutações no PTEN, no TP53, no STK11 e no ATM (Thull & Vogel, 2004). As características clínicas sugestivas de predisposição para cancro hereditário da mama incluem: o aparecimento precoce (antes dos 45 anos de idade), a bilateralidade e a história familiar de cancro da mama ou ovário (mais de 3 familiares com cancro da mama e ovário, pelo menos 2 familiares em primeiro grau afectados com cancro da mama ou cancro da mama masculino).

Espera-se que a análise mutacional de uma grande variedade de genes permita identificar factores importantes na patogénese do cancro da mama, nomeadamente, os genes envolvidos, a natureza das interacções entre esses genes, a interacção com factores ambientais e o eventual papel de alelos protectores.

## Padrões de expressão de genes

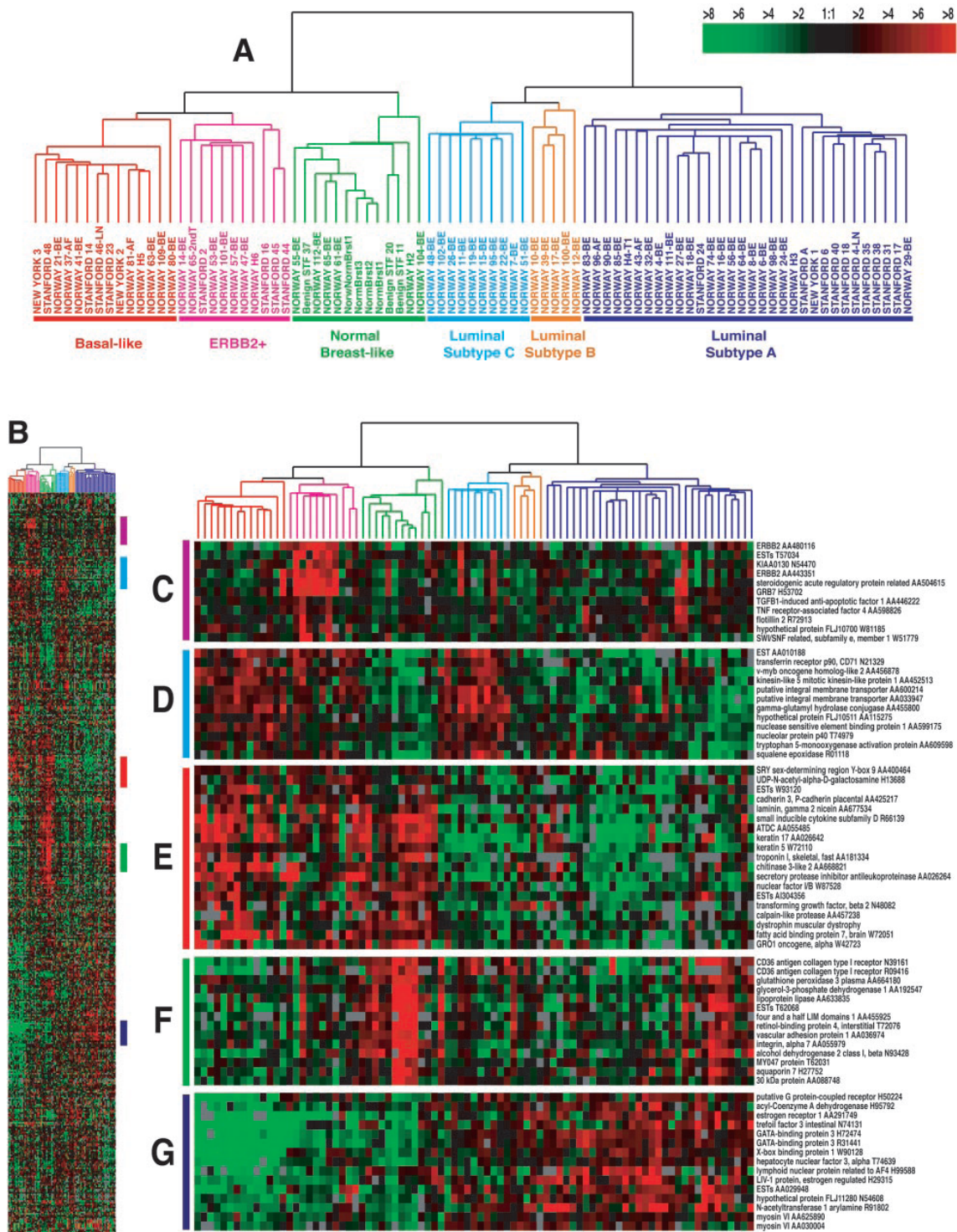
Os estudos de microarrays de cDNA permitem distinguir a alteração dos padrões globais de expressão de genes que ocorre nas células neoplásicas, o que pode ajudar a compreender o comportamento de certas neoplasias. Os padrões de expressão de genes revelam “impressões moleculares”, que são distintas para cada tumor e que refletem propriedades inerentes ao tumor (Eisen *et al*, 1998). A compreensão da relação entre padrões de expressão de genes e a evolução clínica de um tumor é um elemento fundamental para se compreenderem fenótipos como a resistência ou sensibilidade a determinadas terapias, a invasibilidade ou o potencial metastático.

Os resultados obtidos a partir da análise por microarrays mostram que os tumores da mama (**Figura 15**) podem ser, tal como clinicamente descrito, incluídos em dois ramos principais – positivos (ER+) ou negativos (ER-) para receptores de estrogénio (ER). Uma das implicações relevantes dos estudos de microarrays é a divisão dos carcinomas da mama clinicamente designados como “receptores de estrogénio negativos”, em subtipos tumorais biologicamente distintos: basal, ErBB2+ e normal (Perou *et al*, 2000). Assim, um grupo de tumores considerados na clínica como entidades semelhantes poderá incluir neoplasias distintas, o que pode ter consequências a nível de prognóstico, diagnóstico ou terapêutico.

O subtipo basal, o *Erb-B2*+ e o normal caracterizam-se pela ausência ou reduzida expressão de ER e de vários factores de transcrição expressos pelas amostras contidas no subtipo luminal/ ER+.

Os carcinomas de tipo basal distinguem-se pela expressão de queratinas tipicamente expressas por células mioepiteliais ou por potenciais células progenitoras da mama. O subtipo basal é caracterizado pela elevada expressão de queratinas 5 e 17, laminina e proteína-7 de ligação a ácidos gordos. Além da expressão de queratinas específicas, os carcinomas basais apresentam em comum com as células mioepiteliais vários genes associados com a proliferação e outros como a caderina-P. Os carcinomas causados por mutações no gene BRCA1 também se agrupam no subtipo basal. Algumas das características que os carcinomas causados por mutações no gene BRCA1 partilham com os carcinomas de tipo basal são: a ausência de expressão de ER e de HER2, a expressão de queratinas de tipo basal e a reduzida diferenciação. Apesar desta partilha

de algumas características moleculares, a maioria dos carcinomas tipo basal não apresentam mutações no gene BRCA1.



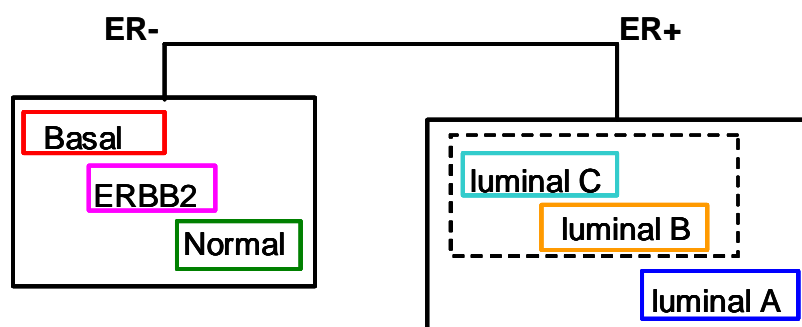
**Figura 15:** Padrão de expressão de genes de 85 amostras experimentais que representam 78 carcinomas, 3 tumores benignos e 4 tecidos normais, analisados por *clustering* hierárquico.

(Sorlie *et al*, 2001)

O subtipo *Erb-B2+* é caracterizado pela elevada expressão de vários genes do amplicão 17q22.24, incluindo o oncogene ERBB2 e outros genes adjacentes.

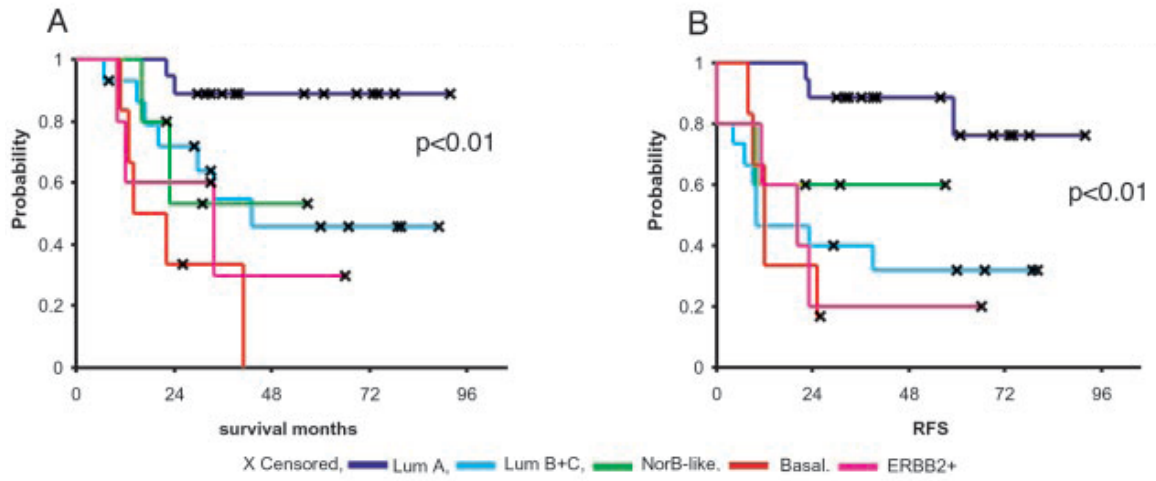
O “subtipo normal” é caracterizado pela mais elevada expressão de vários genes conhecidos por serem expressos pelo tecido adiposo e por outras células de tipo não epitelial. Os tumores incluídos neste grupo também apresentam uma elevada expressão de genes característicos das células mioepiteliais e uma reduzida expressão de genes característicos das células luminais.

Os tumores incluídos no tipo ER+ caracterizam-se pela expressão relativamente elevada de muitos genes que se encontram expressos nas células luminais mamárias. A extensão do número de amostras analisadas permitiu a separação do grupo caracterizado como luminal/ ER+ em dois ou possivelmente três subgrupos (**Figura 16**), que apresentam uma correlação com parâmetros clinicamente relevantes (Sorlie *et al*, 2001).



**Figura 16:** Representação esquemática da divisão dos tumores da mama, com base nos padrões de expressão génica.

Os subtipos tumorais identificados representam grupos de pacientes clinicamente distintos, nomeadamente em relação aos tempos de sobrevida total e de sobrevida sem recidivas (**Figura 17 A e B**) (Sorlie *et al*, 2001). Verificou-se que os subtipos tumorais basal e ERBB2+ estão associados a um menor tempo de sobrevida assim como de sobrevida sem remissões. O subtipo basal pode representar uma entidade clínica que está associada com uma elevada frequência de mutações no gene TP53. Estudos prévios mostraram que as mutações no gene TP53 estão associadas com uma baixa resposta a terapia sistémica e a um mau prognóstico (Berns *et al*, 2000). Observou-se que o comportamento dos tumores classificados como luminal A ou luminal B+C é diferente, com o segundo subtipo a poder representar um grupo clinicamente distinto, associado a um pior prognóstico, concretamente no que diz respeito a recidivas.



**Figura 17:** Análise da sobrevivência e sobrevida sem recidivas (RFS: relapse-free survival), de 49 pacientes diagnosticados com cancro da mama, baseada na classificação das diferenças dos padrões de expressão de genes.

(Sorlie *et al*, 2001)

## **Estadiamento e Prognóstico**

O grau de estadiamento representa o estado da doença ou potencial biológico do tumor de um paciente individual. O estadiamento baseia-se sobretudo em informações clínicas e patológicas. Um estadiamento rigoroso permite um bom prognóstico, planear a estratégia terapêutica mais adequada ao paciente e comparar resultados terapêuticos.

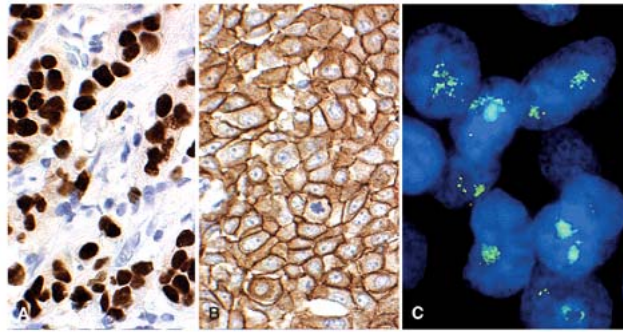
O prognóstico é determinado através do exame patológico do carcinoma primário e dos nodos linfáticos axilares, excepto nos casos em que no momento da detecção já existem metástases ou carcinoma inflamatório. A classificação TNM, concebida pela União Internacional Contra o Cancro (UICC), é um padrão de classificação mundialmente aceite, que caracteriza um tumor pela extensão anatómica. O sistema TMN baseia-se nas características clínicas do Tumor ( T ), nos nodos linfáticos regionais ( N ), e na presença ou ausência de metástases distantes ( M ).

Na ausência de metástases distantes, o estado do nodo linfático axilar é o factor de prognóstico mais importante para o carcinoma invasivo, sendo um indicador específico da agressividade biológica das células neoplásicas ou da incapacidade de defesa do hospedeiro. Tradicionalmente, a avaliação clínica de envolvimento axilar consiste na dissecação axilar completa e posterior análise de uma secção de cada um dos nodos linfáticos removidos, corados com hematoxilina e eosina. Uma vez que a dissecação axilar completa implica uma elevada morbilidade e que a remoção de nodos histologicamente negativos não apresenta qualquer benefício, tem sido cada vez mais comum o uso da Biópsia do Nodo Sentinela. A Biópsia do Nodo Sentinela é um procedimento cirúrgico minimamente invasivo, que se baseia no facto de a maioria dos carcinomas drenarem para o designado Nodo Sentinela. O estado do gânglio sentinela permite prever com elevada segurança o estado dos gânglios linfáticos remanescentes. Com esta técnica apenas os nodos linfáticos histologicamente positivos são submetidos a análise, o que permite além de uma grande redução da morbilidade associada à dissecação axilar completa, uma análise patológica mais minuciosa e menos dispendiosa. De acordo com os principais factores de prognóstico, os carcinomas da mama são divididos nos estádios clínicos 0, I, II, III, IV, que são identificados de acordo com as seguintes características:

- Estádio 0: CDIS ou CLIS. A taxa de sobrevivência de 5 anos é de aproximadamente 92%.
- Estádio I: carcinoma invasivo com diâmetro inferior ou igual a 2 cm e sem envolvimento nodal, ou apenas com metástases inferiores a 0.02 cm de diâmetro; também inclui carcinoma *in situ* com micro-invasão. A taxa de sobrevivência de 5 anos é de aproximadamente 87%.
- Estádio II: carcinoma invasivo com diâmetro inferior ou igual a 5 cm, com envolvimento nodal que pode ir até três nodos, ou com diâmetro superior a 5 cm mas sem envolvimento nodal. A taxa de sobrevivência de 5 anos é de aproximadamente 75%.
- Estádio III: carcinoma invasivo com diâmetro inferior ou igual a 5 cm, com envolvimento nodal de um número acima de 4 nodos linfáticos; carcinoma invasivo com diâmetro superior a 5 cm, sem envolvimento nodal; carcinoma invasivo com 10 ou mais nodos envolvidos; carcinoma invasivo com envolvimento do nodo linfático mamário ipsilateral interno; ou carcinoma invasivo com envolvimento de pele; fixação da parede do peito; carcinoma inflamatório. A taxa de sobrevivência de 5 anos é de aproximadamente 46%.
- Estádio IV: qualquer cancro da mama com metástases distantes. A taxa de sobrevivência de 5 anos é de aproximadamente 13%.

A aplicação da mamografia, como técnica de rastreio de cancro da mama, levou a um aumento considerável da detecção de carcinomas ductais *in situ* ou carcinomas invasivos pequenos. Os CDIS não são detectados por palpação nem por inspecção visual, mas apresentam-se frequentemente como calcificações mamográficas. Actualmente, 15 a 30% de todos os carcinomas detectados por mamografia são casos de CDIS. A mastectomia é um tratamento que dá uma probabilidade de cura de 95% dos casos de CDIS. A conservação da mama é um método associado a uma muito menor morbidade e é apropriada para a maioria das mulheres com CDIS, embora possa resultar num risco ligeiramente mais elevado de recorrência. Mesmo quando há preservação da mama, a morte por cancro da mama, de mulheres com CDIS adequadamente tratadas, é muito rara. A rara recorrência ou morte deve-se a DCIS residual nos ductos, no tecido adiposo subcutâneo que não foi removido durante a cirurgia ou a focos de invasão que não foram detectados no diagnóstico.

Há factores de prognóstico que, embora não sejam considerados principais, permitem prever o comportamento do tumor a terapias específicas. Geralmente, esses factores são marcadores moleculares produzidos pelas células tumorais ou pelo hospedeiro (**Figura 18**) e estão correlacionados com a actividade biológica tumoral.



**Figura 18:** Factores de prognóstico. A: O receptor de estrogénio é detectado no núcleo por estudos imuno-histo-químicos. O receptor de progesterona apresenta a mesma aparência. B: A sobre-expressão de Her2/neu é detectada na membrana celular por imuno-histo-química. A amplificação do gene *Her2/neu* pode ser detectada com recurso à análise por FISH.

(Lester SC, 2005)

#### - Os receptores de estrogénio e de progesterona:

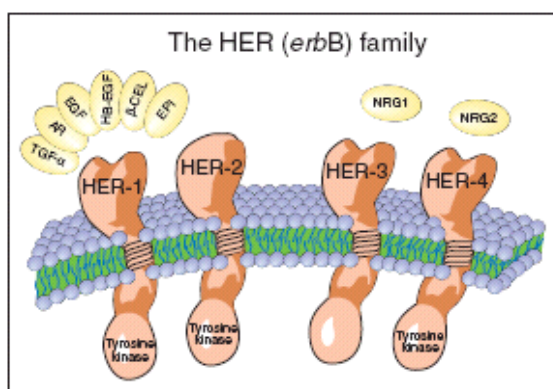
A avaliação dos receptores de hormonas é um dos factores mais valiosos para prever a resposta à terapia hormonal. O prognóstico para cancros positivos para receptores de hormonas é ligeiramente melhor do que para os tumores negativos. Os tumores positivos para ER são mais comuns em mulheres pós-menopausa e estima-se que entre 70-80% dos carcinomas da mama expressam esses receptores. A probabilidade de tumores com receptores de estrogénio e progesterona responderem a manipulação hormonal é de 80%, declinando para 40% quando apenas um tipo de receptor responde à terapia hormonal. A terapia endócrina mais utilizada no tratamento de cancro da mama é o tamoxifeno, que é uma droga que compete com o estrogénio na ligação ao seu receptor.

Os padrões de expressão de genes podem ser úteis para interpretar casos de cancros que expressam ER e que não respondem a tratamentos hormonais. A ausência de resposta

aos tratamentos hormonais pode ser uma consequência da disrupção da via de sinalização, o que nos padrões de expressão é sugerido pela baixa expressão de genes regulados por ER. De facto, o padrão de expressão de genes regulados por ER permite uma melhor previsão do comportamento tumoral e da resposta a agentes bloqueadores de estrogénio do que a expressão do ER por si só.

### - HER2/neu:

O proto-oncogene HER2/neu/ c-erbB-2 (Human Epidermal Growth Factor Receptor 2) está localizado no braço longo do cromossoma 17 (17q21-22) e codifica uma glicoproteína com actividade de tirosina-cínase, da família HER ou EGFR (Epidermal Growth Factor Receptor) (Slamon *et al*, 1987) (**Figura 19**).



### Legenda:

- TGF-α:** Transforming Growth Factor α
- AR:** Amphiregulin
- EGF:** Epidermal Growth Factor Receptor
- HB-EGF:** Heparin Binding EGF
- β-cell:** β cellulin
- EP1:** EPinephrine
- NRG1 e NRG2:** NeuReGulin 1 e 2

**Figura 19:** A família HER.

(Ross *et al*, 2003)

A amplificação ou sobre-expressão do c-erbB-2 ocorre em 20-30% dos cancros da mama e parece estar associada a um fenótipo mais agressivo, com diminuição da sobrevida livre de doença e sobrevida total. Na maioria dos carcinomas, a amplificação do HER2/neu tem como consequência o aumento da transcrição em mRNA e a sobre-expressão da proteína. O HER2 parece actuar como co-receptor para múltiplos factores de crescimento. Após a ligação de um ligando ao domínio extra-celular do HER2, ocorrem uma série de alterações que promovem a formação de dímeros. Pensa-se que a formação de dímeros regula a actividade tirosina-cínase e a auto-fosforilação. A activação do receptor vai levar à fosforilação de proteínas intra-celulares, que actuam como mensageiros secundários. A via de transdução de sinal desencadeada depende do ligando e dos receptores que dimerizam com o HER2 (Ross *et al*, 2003).

Existe, desde 1998 um anticorpo monoclonal – conhecido por trastuzumab ou pelo nome comercial Herceptin® – que tem como alvo as células tumorais que sobre-expressam o HER2 e que se liga ao domínio extra-celular do receptor e funciona como um inibidor da sua activação. Este é um exemplo paradigmático de como os avanços de conhecimento em investigação básica se podem traduzir em informações clínicas, concretamente na detecção de marcadores tumorais relevantes, que também servem como alvos terapêuticos.

Uma vez que uma célula epitelial típica leva cerca de 20 a 30 anos a acumular alterações genéticas para progredir para doença metastática, as mulheres deverão informar-se adequadamente sobre as várias, e por vezes equivalentes, opções terapêuticas. Para casos de cancro da mama mais avançados, com metástases distantes, o objectivo deverá ser proporcionar à mulher um prolongamento da sobrevida associado a uma redução da morbilidade (Lester SC, 2005).

## Mecanismos de tumorigénese

A transformação das células epiteliais normais da mama em células neoplásicas ocorre em várias etapas (Dahiya R & Deng G, 1998; Brenner AJ & Aldaz CM, 1997). As alterações morfológicas associadas com risco reduzido de cancro da mama são lesões em que há um aumento do número de células epiteliais – alterações proliferativas. Neste estado precoce é visível uma expressão anómala de receptores hormonais e a consequente desregulação da proliferação. Foi sugerido que as alterações precoces estão relacionadas com insensibilidade a sinais inibidores do crescimento, auto-suficiência em sinais de crescimento e evasão à apoptose. Por exemplo, mutações nos genes ER, EGFR, RAS e HER2/neu podem resultar na auto-suficiência em sinais de crescimento. Por outro lado, as alterações em genes como o p53, que tem um papel fulcral no controlo do ciclo celular, na reparação de DNA e na apoptose, podem implicar a alteração de várias funções biológicas.

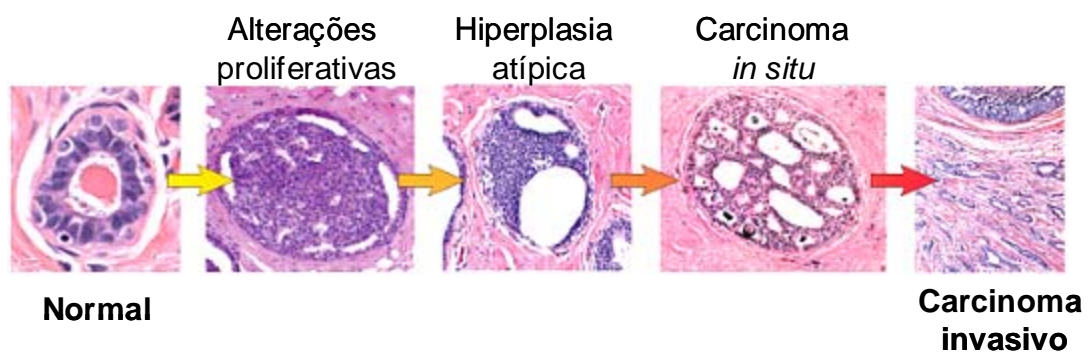
A perda de heterozigotia (*LOH*) parece ser uma alteração tardia, já que raramente é detectada nas alterações proliferativas, é frequente nas hiperplasias atípicas e está sempre presente nos carcinomas *in situ*.

Foi proposto que a capacidade de metastização resulta de uma anomalia genética que ocorre logo no início do processo neoplásico, porém esta hipótese não foi comprovada e este é um dos processos ainda mal esclarecidos. Alguns dos processos biológicos que têm sido sujeitos a intensa investigação são os envolvidos nas fases mais precoces da metastização, como o aumento de mobilidade das células e produção de proteases que degradam a matriz extra-celular.

O vasto conjunto de alterações histológicas (**Figura 20**) são uma manifestação exterior visível das alterações biológicas que ocorreram nas lesões e nos carcinomas da mama e sugerem que existem várias vias de tumorigénese.

A mama normal mantém a sua integridade graças a uma série de interacções complexas entre as células luminais, as mioepiteliais e as do estroma (Wiseman BS & Werb Z, 2002). Além das alterações acumuladas nas células luminais, têm que ocorrer alterações paralelas nas células mioepiteliais e nas do estroma. O potencial maligno é devido à combinação de eventos genéticos, epigenéticos ou à disrupção de vias de sinalização

inter-celulares normais (Tlsty TD & Hein PW, 2001). A perda da integridade da membrana basal, causada por anomalias nas funções das células do estroma e mioepiteliais, é uma das possibilidades que explica a transição de carcinoma *in situ* para carcinoma invasivo (Barsky SH & Karlin NJ, 2005). Em certos tipos de CDIS é evidente um aumento de angiogénese à volta da membrana basal de certos ductos. Esse aumento pode dever-se à estimulação directa das células malignas, aos efeitos secundários das células do estroma ou à perda da capacidade das células mioepiteliais para inibirem a angiogénese (Lester SC, 2005).



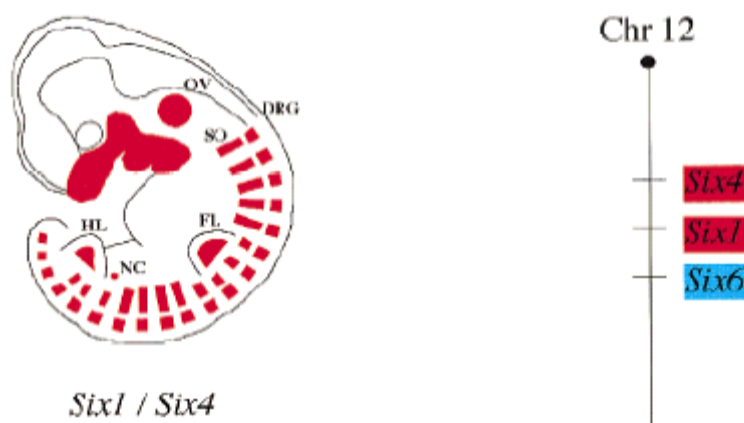
**Figura 20:** Alterações morfológicas dispostas de acordo com o risco subsequente para carcinoma invasivo.

(adaptado de Robbins & Cotran *in* Pathologic Basis of Disease)

## Genes SIX e carcinoma da mama

A variação nos programas de transcrição é um dos principais factores que influenciam a diversidade biológica das células humanas. O SIX4 é um factor de transcrição no qual foram encontradas mutações em carcinoma da mama, pelo que é uma hipótese racional que este gene possa ter um papel no processo de tumorigénese da mama.

O SIX4 é um gene ainda relativamente mal estudado e a literatura que existe, relativamente escassa, refere-se sobretudo ao seu papel na embriogénese (Ozaki *et al*, 2001; Kobayashi *et al*, 2007) Na tentativa de esclarecer as funções do Six4 foram criados ratos deficientes em Six4 (Ozaki *et al*, 2001). Após o nascimento, os ratos apresentavam uma aparência anatómica normal. Os autores propuseram que a perda da função do Six4 possa ser compensada pela expressão normal de Six5 ou Six1, dado que, o Six5 apresenta uma arquitectura molecular muito semelhante e, além dos padrões de expressão espacial do Six1 e Six4 serem idênticos (Oliver, 1995) (**Figura 21**), a sua localização em *loci* muito próximos, sugere que elementos de regulação sejam partilhados e que os genes alvo sejam comuns.



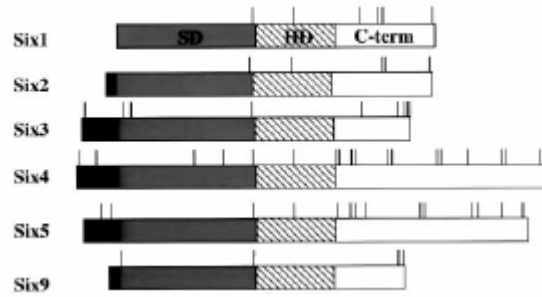
**Figura 21: Representação esquemática dos padrões de expressão embrionários e localização cromossômica do Six1 e Six4, no rato.** Ov, otic vesicle; DRG, Dorsal root ganglion; SO, somites; FL, forelimb; NC, nephrogenic cord; HD, hindlimb.

(adaptado de Kwakami *et al*, 2000)

O SIX1 é um gene associado à progressão tumoral das neoplasias da mama humana. No tecido normal de adulto humano o HSIX1 não é expresso na mama, mas é expresso em músculos esqueléticos, nas glândulas salivares e pituitária, no pulmão, na traqueia e no rim (Ford *et al*, 1998). Embora nos tecidos mamários normais os níveis de expressão de SIX1 sejam muito reduzidos ou mesmo ausentes, são elevados em cancro da mama primário e atingem um valor máximo nas lesões metastáticas: a sobre-expressão de HSIX1 foi observada em 44% de cancros da mama primários e em 90% de lesões metastáticas (Ford *et al*, 1998).

Durante a interfase e a mitose, a HSIX1 é uma fosfo-proteína nuclear e, pelo menos em parte, a sua fosforilação deve-se à cínase de caseína II (CK2) (Ford *et al*, 2000). A CK2 é uma cínase de serina/ treonina que é ubiquamente expressa, altamente conservada nos organismos eucarióticos (Allende & Allende, 1995) e que fosforila vários factores de transcrição (Kasahara & Izumo, 1999). Na mitose observa-se uma hiperfosforilação da HSIX1 e uma perda concomitante da sua actividade de ligação ao DNA, pelo que se especula que a sua actividade possa ser regulada pelo estado de fosforilação (Ford *et al*, 2000). Esta hipótese é corroborada pelo facto de a regulação por fosforilação da actividade de múltiplas homeoproteínas provocar alterações na actividade de ligação ao DNA, na transactivação e na localização nuclear (Kasahara & Izumo, 1999).

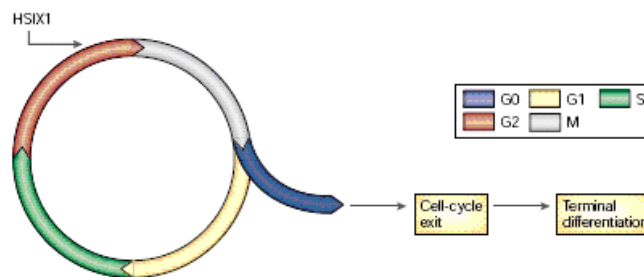
Nos Six, a maioria dos locais de fosforilação específicos da mitose ocorre na sua extremidade carboxílica e um alinhamento de HSIX1 com outros membros da família SIX (**FIGURA 22**), demonstrou que os vários locais potenciais de fosforilação por CK2 são altamente conservados (Ford *et al*, 2000). Um dos locais de fosforilação para o CK2, conservado em Six1, Six2, Six4 e Six5, encontra-se imediatamente adjacente à segunda hélice do homeodomínio, que é importante por fornecer o cerne hidrofóbico do homeodomínio e preservar a natureza anfipática das hélices (Shang *et al*, 1994).



**FIGURA 22:** Diagrama de representação dos vários locais de fosforilação da cínase de caseína II (CK2), das homeoproteínas Six. A linha vertical representa um local putativo CK2. SD, Six domain; HD, Homeodomain; C-term, C-terminal region.

(Ford *et al*, 2000)

Os modelos de cultura de células sugerem que a actividade oncogénica do SIX1 é uma consequência da ausência de regulação do controlo do ciclo celular, verificando-se que a sobre-expressão do gene SIX1 permite escapar ao ponto de controlo (*checkpoint*) G2 do ciclo celular, após exposição a raios X. Foi proposto (Ford *et al*, 2000) que, durante o ciclo celular, aquando da transição G2/M, a actividade do HSIX1 (**Figura 23**) é regulada pela CK2 e que a desregulação da via em que ambas as moléculas participam leva à progressão tumoral.

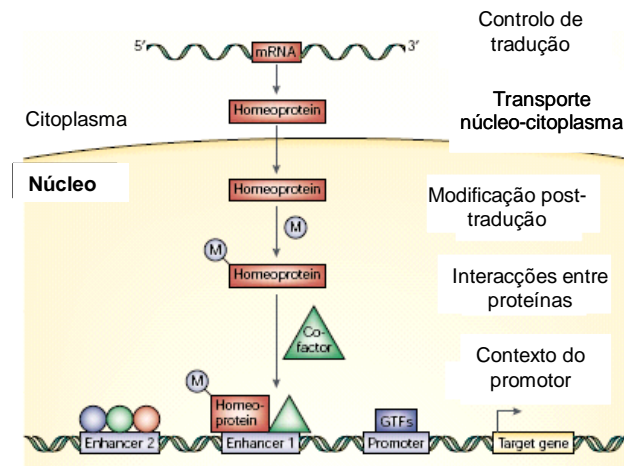


**Figura 23:** Envolvimento do HXIS1 no ciclo celular

(Abate-Shen C, 2002)

Para compreender melhor a função dos genes Six é necessário investigar quais os seus genes alvo, os genes de regulação e as moléculas de sinalização que modificam a expressão e função das proteínas Six bem como os genes que estão envolvidos no controlo do crescimento celular e na adesão e movimento. Em geral, há poucos dados sobre os genes alvos das homeoproteínas e, apesar de *in vitro*, a maioria dos

homeodomínios se ligarem a pequenas sequências de DNA semelhantes, *in vivo* apresentam funções muito específicas. Acredita-se que a especificidade funcional das homeoproteínas é regulada por vários mecanismos, como o controlo *post*-transcricional, a translocação nuclear e interações proteína-proteína ou com co-factores (Kasahara & Izumo, 1999 (**Figura 24**)).



**Figura 24:** Mecanismos de regulação da actividade das homeoproteínas.  
(Abate-Shen C, 2002)

# **Metodologia**

### ➤ **Recolha de amostras de carcinoma ductal**

Todas as amostras (**Tabela 2**) foram obtidas de acordo com a *Health Insurance Portability and Accountability Act* (HIPAA).

Os tumores primários de mama, e tecido cirúrgico normal circundante, foram isolados a partir de pacientes de sexo feminino, diagnosticadas com carcinoma ductal, e obtidos através do Banco de Tecido do Centro de Cancro (*Cancer Center Tissue Bank*) da Carolina do Sul, Estados Unidos da América.

Cada amostra de tecido foi congelada no máximo após 30 minutos da excisão e armazenada a -80°C.

### ➤ **Micro-dissecção**

As secções de 20 µm de tecido tumoral primário da mama foram colocadas em *Sigma silane-prep™ slides* e coradas com hematoxilina e eosina. Seguindo um protocolo padrão, separaram-se as células epiteliais malignas do tecido circundante morfológicamente normal, e por micro-dissecção de “captura com laser” (PixCell<sup>®</sup> Ite, Arcturus), recuperou-se cada população de células.

O DNA genómico (gDNA) foi extraído e purificado, a partir de 20 lâminas para cada amostra, com o *Kit QIAamp DNA Micro*, da Qiagen<sup>™</sup>, de acordo com o protocolo do fabricante.

### ➤ **Amplificação do genoma completo**

A amplificação do genoma completo (*Whole genome Amplification*) consistiu na desnaturação do DNA molde com 5M de hidróxido de potássio (KOH), posterior neutralização e incubação a 30°C, durante 16 a 24 horas, com tampão REPLI-g e REPLI-g polimerase de DNA, de acordo com o protocolo do fabricante, seguida de inactivação da enzima por incubação das amostras a 65°C, durante 3 minutos, e armazenamento a -20°C.

Número de casos	Designação oficial	T	N	M
1	23876	2	0	0
2	29161	1A	0	0
3	29716	2	0	0
4	31498	1	0	0
5	10043	1C	1A	0
6	25173	2	1B	X
7	12201	1	0	0
8	23659	1C	0	X
9	29419	2	0	0
10	30262	1C	0	0
11	40186	2	0	0
12	40422	1C	0	1
13	40626	2	0	X
14	40635	1C	0	0
15	40731	2	0	0
16	4662	1C	1A	0
17	5426	2	3	0
18	18848	1C	1	0
19	40130	2	1	0
20	10981	2	1A	0
21	11010	1B	3	0
22	11936	1C	2A	0
23	29119	1C	1A	0
24	31196	1C	1A	X
25	31415	2	2A	0
26	19689	1C	1A	0
27	22880	1B	1A	X
28	29374	2	2A	X
29	31108	1C	1M	X
30	29223	1	1A	0
31	29693	2	3A	0
32	29133	2	1	0
33	40172	4B	0	0
34	40312	2	0	0
35	12871	2	0	0
36	18738	2	0	X
37	23909	2	0	0
38	31379	3	0	0
39	40143	1C	0	0
40	10696	1C	1A	0
41	40181	2	2A	0
42	40368	2	1A	0
43	40372	2	1A	0
44	19157	1C	0	0
45	4421	2	1A	0
46	40578	1C	X	X
47	6633	1C	1A	0
48	3550			

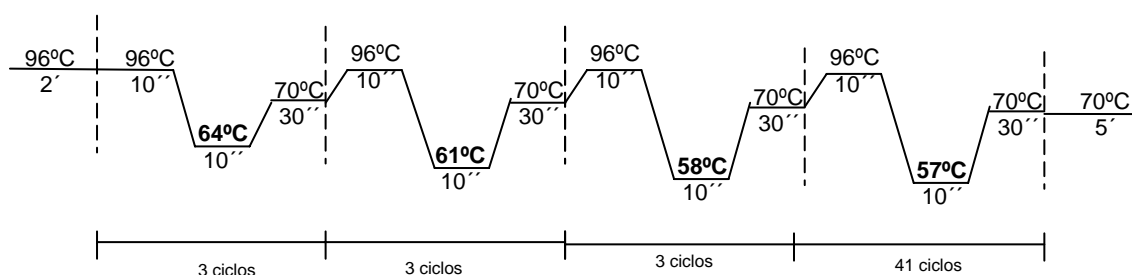
**Tabela 2:** Número, designação oficial e classificação TNM das amostras usadas na busca de alterações somáticas no SIX4.

➤ **Amplificação por PCR da região codificante do SIX4**

Os fragmentos de DNA genómico, cobrindo os exões 1 a 3, do SIX4, foram amplificados por PCR, usando os pares de oligómeros indicados na **Tabela I, em anexo**.

A mistura reacção de amplificação continha um tampão (67mM Tris-HCL, pH8.8, 6.7 mM MCl<sub>2</sub>, 16.6 mM NH<sub>4</sub>SO<sub>4</sub>, 10 mM 2-mercapto-etanol), 1 mM dNTPs (Invitrogen, San Diego, CA), 1µM de cada oligómero (*forward* e *reverse*), 6% DMSO, 2mM ATP, 0.25 U Platinum *Taq* (Invitrogen, San Diego, CA) e 1 ng de DNA.

O protocolo de amplificação, conhecido como *touchdown*, incluiu um passo inicial de 2 min a 96°C, seguido de 3 ciclos de 10s a 96°C, 10s a 64°C, 30s a 70°C, 3 ciclos de 10s a 96°C, 10s a 61°C, 30s a 70°C, 3 ciclos de 10s a 96°C, 10s a 58°C, 30s a 70°C, 41 ciclos de 10s a 96°C, 10s a 57°C, 30s a 70°C, um ciclo de 5min a 70°C e arrefecimento a 4°C.

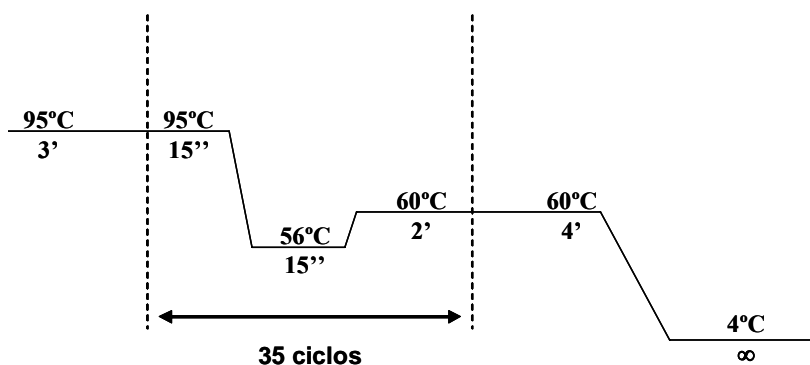


**Figura 25:** Representação esquemática da reacção de amplificação por PCR, dos amplicões dos 3 exões do gene SIX4.

Os produtos de amplificação de ADN genómico foram purificados com o kit AMPure (Agencourt Biosciences, Beverly, MA)

➤ **Reacção de sequenciação**

A pesquisa de mutações foi efectuada por sequenciação dos produtos de PCR, purificados, com o *kit* BigDye v3.1, da Applied BioSystems, com o seguinte protocolo: Incubação inicial a 95°C durante 3 minutos, seguida de 35 ciclos de 15s a 95°C, 15s a 56°C e 2min a 60°C, de uma incubação final de 4min a de 60°C e de arrefecimento a 4°C. A confirmação de sequência de vectores construídos também foi efectuada por sequenciação.



**Figura 26:** Representação esquemática da reacção de sequenciação dos amplicões dos 3 exões do gene SIX4.

#### ➤ **Purificação e Análise dos produtos**

Os produtos resultantes da reacção de sequenciação foram purificados por filtração por coluna de Sephadex-G50 Fine, misturados com 10 µl de formamida (AppliedBioSystems) e analisados num sequenciador ABI prism 3100.

#### ➤ **Análise dos electroforetogramas**

As sequências obtidas foram analisadas por inspecção visual dos electroforetogramas e com o programa DNASIS v5.0-Demo (Hitachi).

#### ➤ **Preparação de vectores**

#### **Extracção de DNA plasmidico (pDNA) de culturas bacterianas:**

Resumidamente, o processo consistiu na sedimentação da cultura em meio líquido, por centrifugação, durante 1 minuto, a 14000g; seguida pela ressuspensão no tampão 50 mM Tris-HCl pH = 8.0; 10 mM EDTA; 100µg/ml RNase A; e pela adição de uma solução alcalina, 200mM NaOH; SDS 1% – lise alcalina. A solução resultante foi neutralizada com 3.0M acetato de potássio pH 5.5 e a suspensão centrifugada durante 10 minutos, a 14000g.

A fracção sobrenadante foi transferida para um novo tubo e o DNA precipitado pela adição de 0.8 volumes de isopropanol e centrifugação durante 5 minutos, a 14000g. A fracção sobrenadante foi desprezada e o precipitado de DNA *lavado* com 1 ml de etanol 70%, e após a sua *secagem*, ressuspendido num volume adequado de LoTE.

O pDNA extraído foi **quantificado** por comparação de intensidade entre a sua banda e a das bandas de um marcador de peso molecular padrão, num gel 0,8% agarose (Eurobio, Courtaboeuf, França), corado com brometo de etídio.

### Digestão com enzimas de restrição

As digestões de vectores ou fragmentos com as enzimas *BamHI* ou *BglII*, foram efectuadas à temperatura de 37 °C, durante um período de 1h, nas condições indicadas na Tabela 3.

Enzima	Sequência de corte	Fornecedor	Tampão de Reacção	Composição do Tampão de Reacção (10x)
<i>BamHI</i>	5'-G <sup>^</sup> A A T C C-3' 3'-C C T A G <sup>^</sup> G-5'	Takara	Takara K	200 mM Tris-HCl (pH8.5) 100 mM MgCl <sub>2</sub> 10 mM Dithiothreitol 1,000 mM KCl
<i>BglII</i>	5'-A <sup>^</sup> G A T C T-3' 3'-T C T A G <sup>^</sup> A-5'	Takara	Takara H	500 mM Tris-HCl (pH7.5) 100 mM MgCl <sub>2</sub> 10 mM Dithiothreitol 1,000 mM NaCl

**Tabela 3:** Enzimas utilizadas em passos de sub-clonagem.

### Desfosforilação de plasmídeos linearizados:

Os plasmídeos linearizados foram desfosforilados pela Fosfatase Alcalina de Intestino de Vítela (CIAP) (Amersham Biosciences, Uppsala, Suécia), pela adição de 2 µl de CIAP, após a digestão, e incubação à temperatura de 37 °C, durante 30 minutos.

Tipicamente, após uma digestão e desfosforilação, o DNA foi purificado através da **extracção com fenol/clorofórmio/álcool isoamílico (25:24:1)** (USB Corporation, Cleveland, EUA) e precipitado por etanol.

### **Precipitação de DNA por etanol**

O método de precipitação de DNA por etanol consistiu na adição sequencial de 1/3 do volume de uma solução aquosa de acetato de amónia 10M e 3,3 volumes de etanol absoluto, seguida por uma centrifugação, a 14000g, durante 5 minutos. A fracção sobrenadante foi desprezada e no tubo permaneceu apenas o precipitado de DNA. No final o DNA foi lavado com 1 mL de etanol 70% e, após a sua secagem, ressuspendido num volume adequado de LoTE.

### **Purificação a partir de bandas de gel de agarose**

As amostras de DNA foram corridas em gel de agarose a 0,8% e as bandas correspondentes aos fragmentos de interesse cortadas e transferidas para um microtubo, e conservada, durante 20 minutos (no mínimo), a uma temperatura de -80 °C. De seguida foram colocadas num bloco térmico (VWR International, West Chester, EUA) a 42 °C, durante 10 minutos. O DNA foi então extraído da agarose através da passagem, por centrifugação a 4400 g durante 10 minutos, por uma coluna de Sephadex G-50 (Amersham Biosciences, Uppsala, Suécia) e submetido a precipitação por etanol.

O DNA purificado é precipitado com etanol, quantificado e utilizado para clonar o fragmento pretendido.

### **Reacção de Ligação**

As reacções de ligação entre fragmentos com extremidades coesivas foram efectuadas à temperatura de 16 °C, durante um período mínimo de quatro horas.

A cada tubo reaccional foram adicionados 10 ng de DNA do vector, 30 ng de DNA da inserção, 2 µl de tampão 10x T4 DNA ligase (MBI Fermentas, Vilnius, Lituânia) e 1 µl da enzima T4 DNA ligase (MBI Fermentas, Vilnius, Lituânia). O volume final de 20 µl foi atingido através da adição de um volume variável de LoTE.

O DNA ligado, extraído e precipitado, foi ressuspendido em 15µl de LoTE, dos quais 5µl foram utilizados para transformar células electrocompetentes.

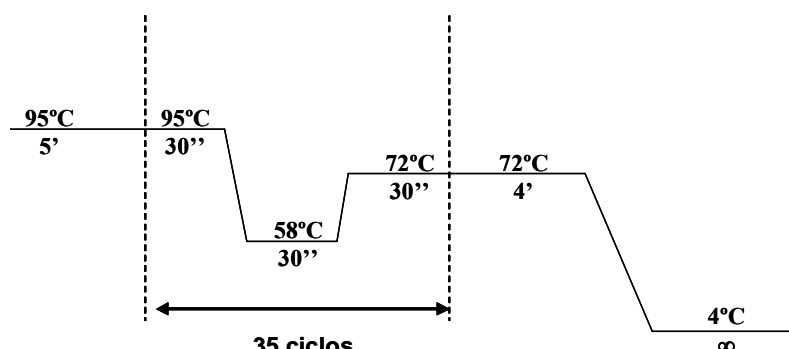
### **Transformação de *E. coli* electrocompetentes**

Os produtos de ligação foram submetidos a extracção por fenol-clorofórmio, precipitados por etanol e ressuspendidos em 15µl de LoTE, dos quais 5µl foram utilizados para transformar células *E. coli* electrocompetentes (produzidas no próprio laboratório) com o auxílio de um electroporador *MicroPulser Electroporator* (Bio-Rad, Richmond, EUA).

### **PCR**

Tipicamente as amplificações por PCR com fins preparativos foram efectuadas usando Foi preparada uma mistura de 25µl contendo tampão de *Taq* DNA Polimerase com (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (MBI Fermentas, Vilnius, Lituânia), 2,5µl dNTPs (2,5 mM cada; QBiogene, Morgan Irvine, EUA), 1,5 µl de MgCl<sub>2</sub> 25 mM (MBI Fermentas, Vilnius, Lituânia) e 175 ng de cada oligonucleotídeo iniciador (Metabion, Martinsried, Alemanha). A essa mistura foi adicionado 1 µl de *Taq* DNA Polimerase (produzida no próprio laboratório) e 1 µl da solução com DNA molde. As reacções de PCR de colónias foram efectuadas usando quantidades proporcionais dos diferentes reagentes por forma a perfazer o volume total de 10µl.

O protocolo de amplificação incluiu um passo inicial de 5 min a 95°C, seguido de um numero variável de ciclos de 30s a 95°C, de 30s a 58°C, de 30s a 72°C, um ciclo de 5min a 70°C e no final um período de arrefecimento a 4°C.



**Figura 27:** Representação esquemática da reacção de amplificação por *PCR* de colónias de bactérias.

### ***PCR* de colónias**

Recolheram-se aleatoriamente colónias individualizadas, com pontas de micropipeta, que foram utilizadas para contactar com uma mistura preparada para reacção de *PCR* e inocular culturas em meio LB (Pronadisa, Madrid, Espanha) com Kanamicina (Eurobio, Courtaboeuf, França).

# **Resultados & Discussão**

## **Análise de alterações somáticas no SIX4**

O objectivo global do projecto de investigação subjacente à tese de dissertação é a clarificação do potencial envolvimento do SIX4 no carcinoma da mama. Assim, para atingir o objectivo proposto, propusemo-nos, na primeira fase da nossa investigação, a pesquisar alterações somáticas no DNA genómico codificante do gene SIX4, proveniente de tumores isolados a partir de pacientes diagnosticadas com carcinoma ductal da mama.

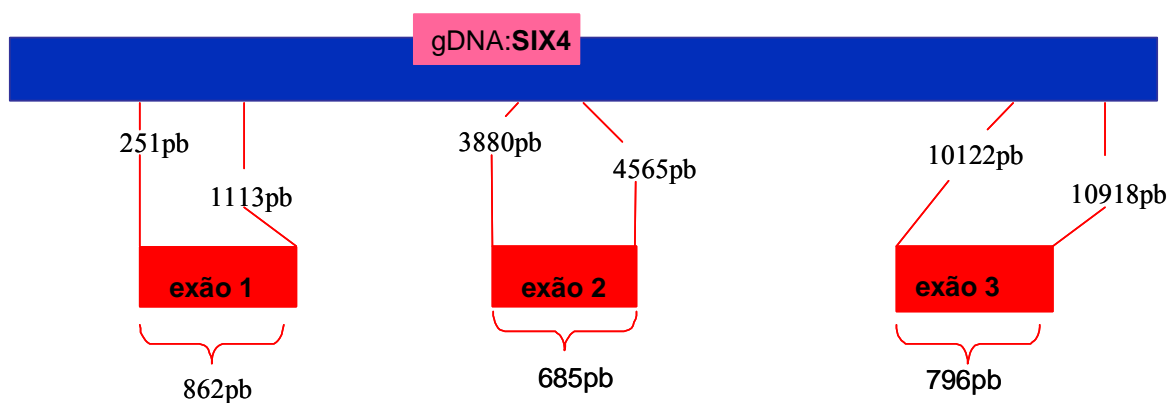
A primeira fase da investigação foi desenvolvida em colaboração com o Prof. Buckhaults, da Universidade de Carolina do Sul, um dos autores do estudo onde foram identificadas mutações somáticas no SIX4 em amostras tumorais e/ ou linhas celulares de carcinoma da mama (Sjöblom T, *et al*, 2006).

As células tumorais foram separadas do tecido normal circundante por micro-dissecção e “captura com laser”. A micro-dissecção e “captura com laser” permite seleccionar uma população de células que exibem a morfologia característica da neoplasia que se pretende estudar e evitar a contaminação com os tecidos adjacentes (Emmert-Buck *et al*, 1996). Assim, o DNA analisado é proveniente de uma população pura de células neoplásicas, o que melhora a sensibilidade da busca de mutações, em particular de amostras tumorais complexas. Além disso, uma vez que o tecido adjacente se mantém intacto, pode utilizar-se o DNA das células adjacentes para análise molecular comparativa.

O DNA genómico extraído dos tecidos “micro-dissectados” foi sujeito a amplificação por *PCR* do genoma completo (*Whole Genome Amplification*). Desta forma, ultrapassou-se uma das limitações geralmente associadas à análise do material *in vivo*, que é a reduzida quantidade de material biológico disponível para investigação.

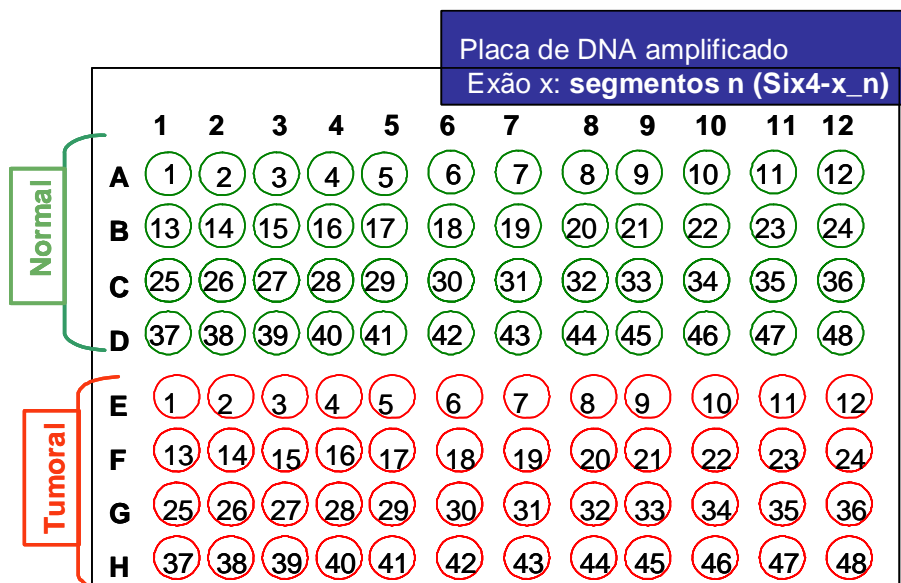
O DNA genómico (sujeito a *Whole Genome Amplification*) foi re-amplificado por *PCR* com oligómeros específicos do SIX4. Em cada par, o oligómero *Forward* tinha, para além da sequência específica do segmento amplificado, uma “cauda” com uma sequência comum, de forma a permitir a utilização do mesmo oligómero (M13) na reacção de sequenciação de todos os amplicões.

O SIX4 é um gene homeobox divergente ou de classe II, que contém três exões (Figura 28).



**Figura 28:** Representação esquemática do gene SIX4 amplificado.

O SIX4 foi amplificado por *PCR*, em reacções paralelas, utilizando gDNA proveniente do tecido tumoral ou do normal circundante (Figura 29).

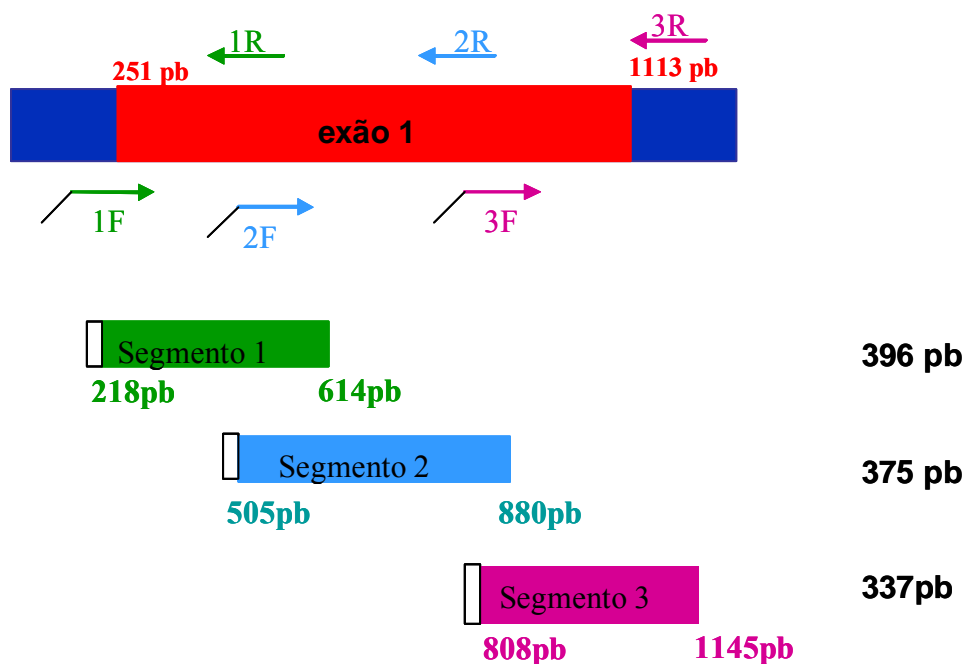


**Figura 29:** Representação esquemática da organização dos produtos resultantes da amplificação por *PCR* do gDNA, extraído a partir do tecido tumoral e normal circundante, dos 48 casos de pacientes diagnosticadas com carcinoma ductal da mama.

Os produtos resultantes foram purificados e os provenientes de tecido tumoral foram sequenciados.

## Pesquisa de alterações somáticas no Exão 1 do SIX4

Para amplificar o exão 1, desenharam-se três pares de pares de oligómeros (Tabela I, em anexo) Figura 30.

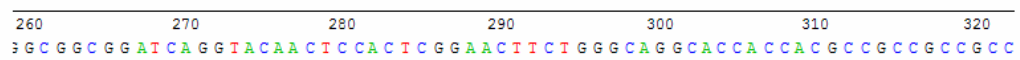
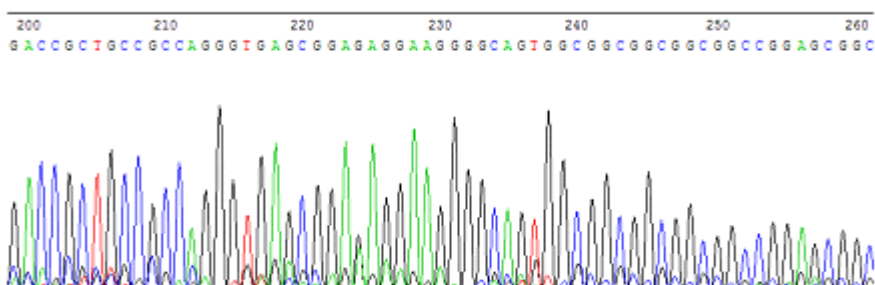
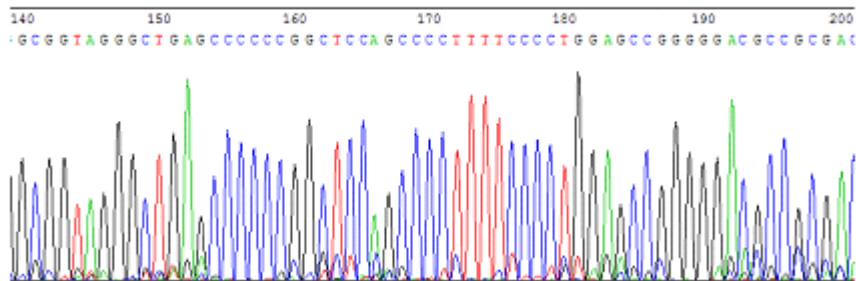
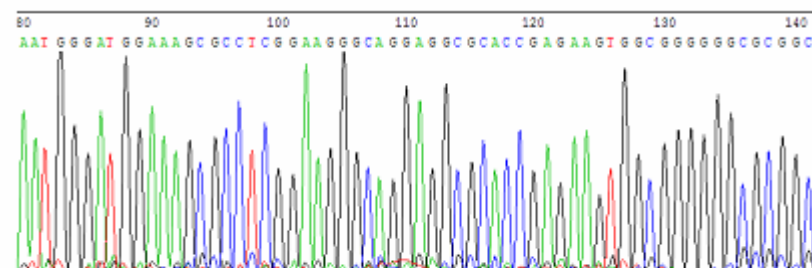
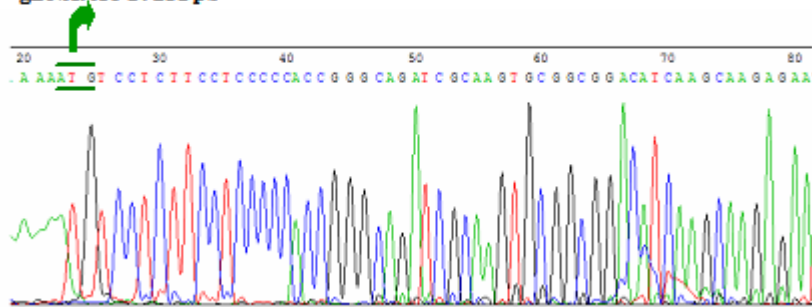


**Figura 30:** Representação esquemática da amplificação do exão 1 do gene SIX4.

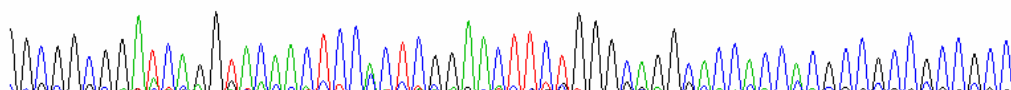
Resultaram 3 amplicões que foram purificados e sequenciados com o oligómero M13.

**Codão de iniciação**

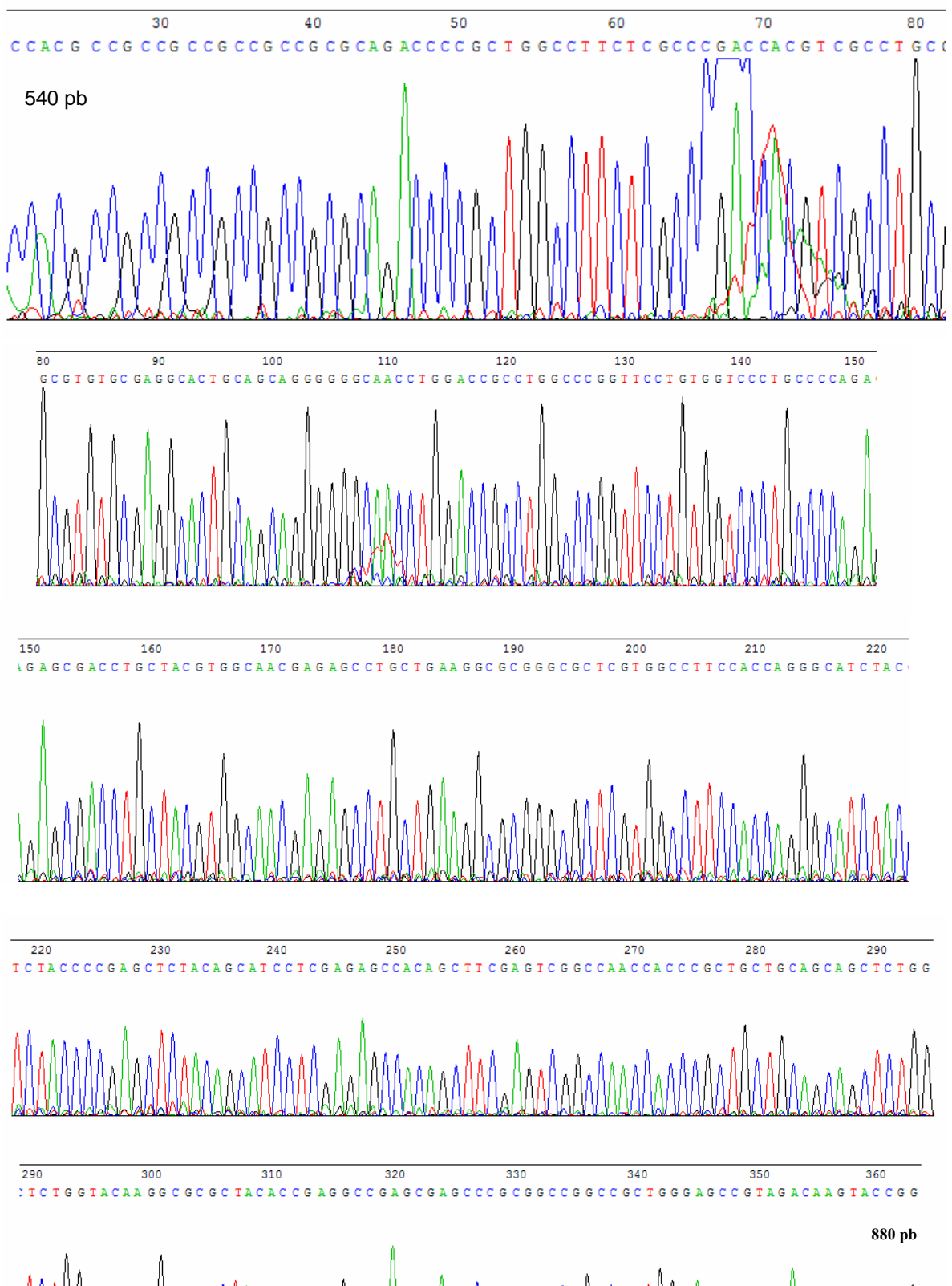
**gDNA: ATG: 251 pb**



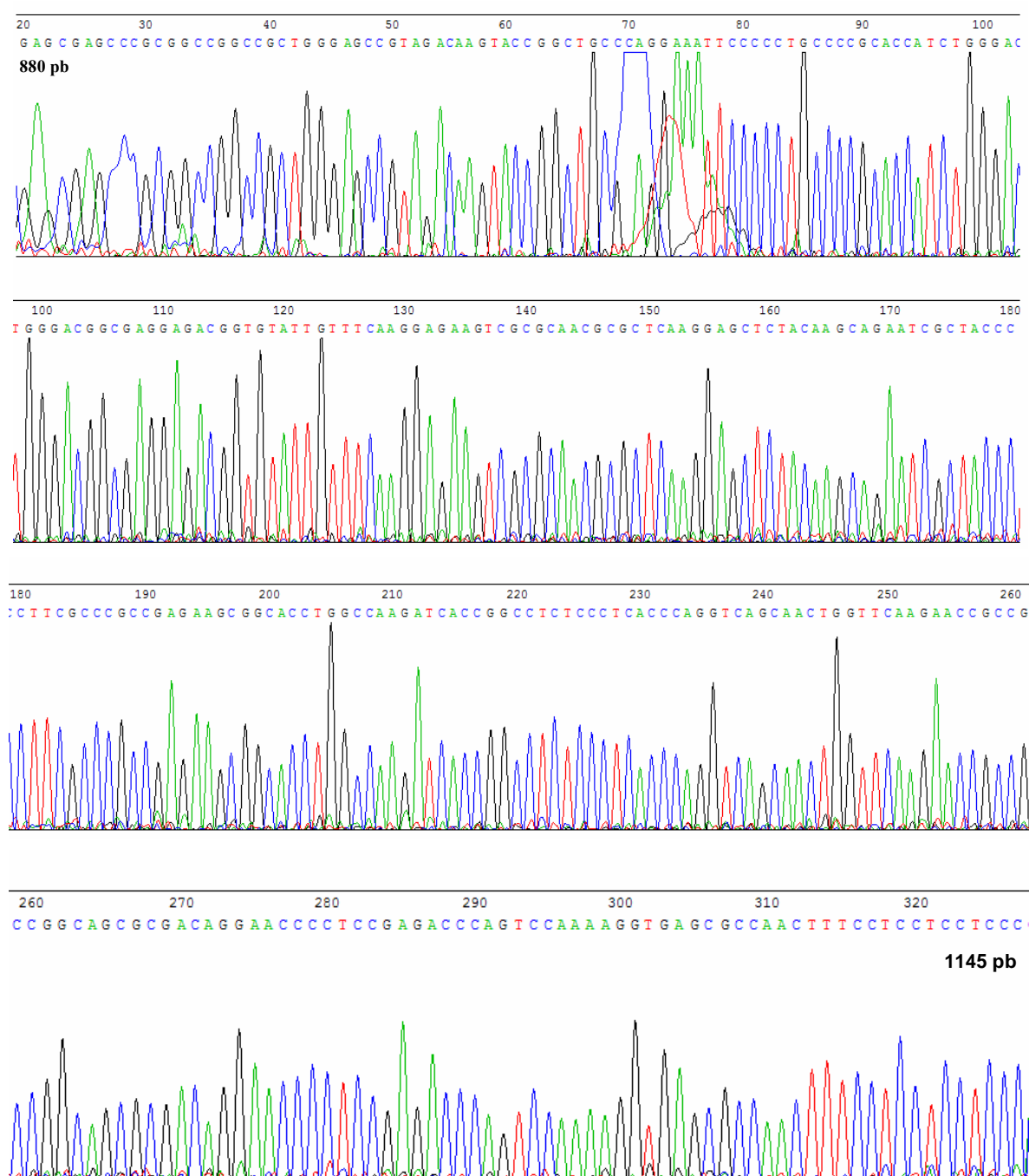
**550 pb**



**Figura 31:** Electroforetograma correspondente à sequenciação do amplicão 1, exão 1, do SIX4.



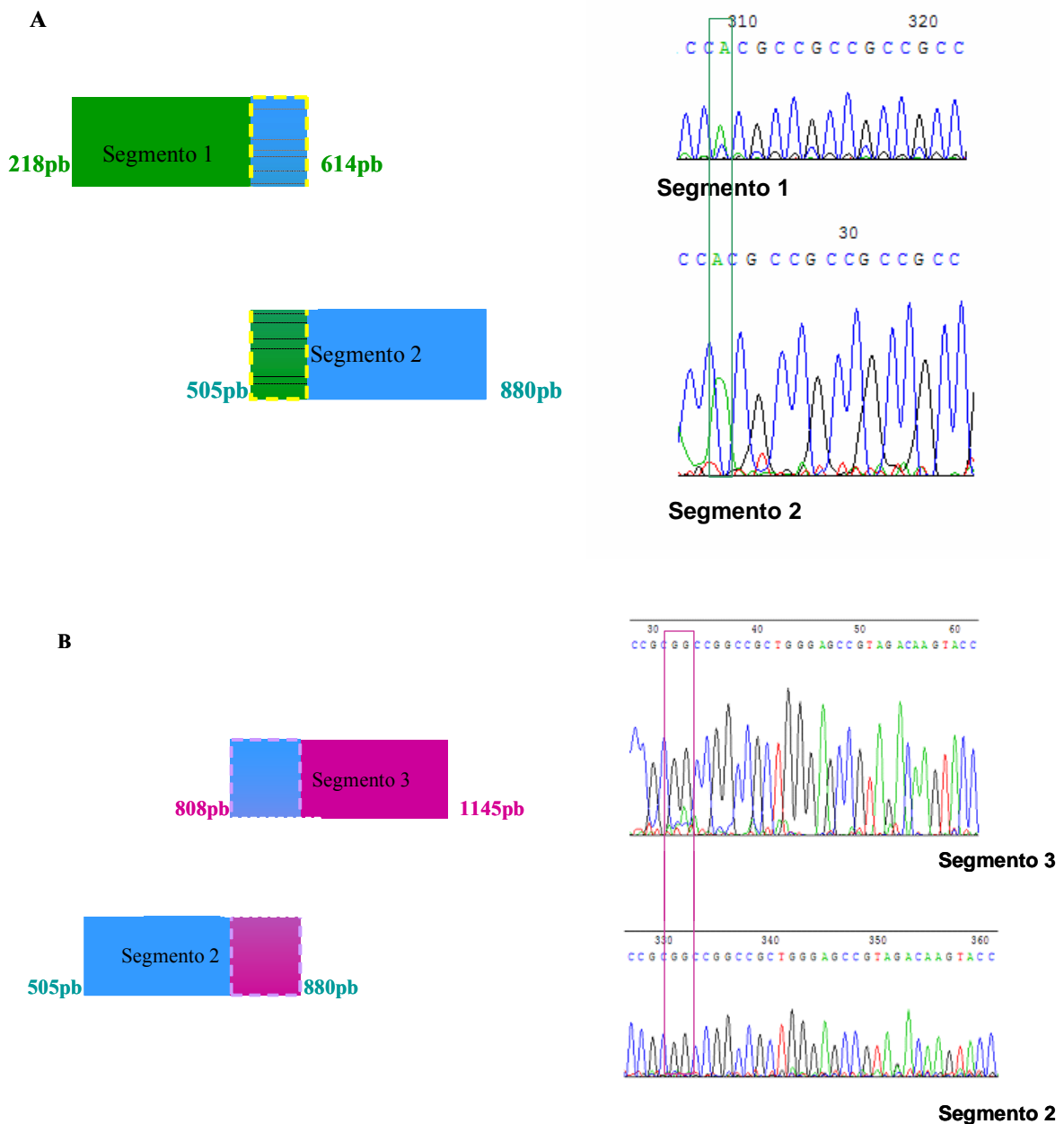
**Figura 32:** Electroforetograma correspondente à sequenciação do amplicão 2, exão 1, do SIX4.



**Figura 33:** Electroforetograma correspondente à sequenciação do amplicão 3, exão 1, do SIX4.

Verificou-se, como ilustrado na **Figura 34 A e B**, que o sinal dos últimos pares de bases sequenciados era relativamente baixo e que os primeiros apresentavam ruído, o que por vezes levou a que se suspeitasse da existência de alterações somáticas.

A comparação da região de sobreposição dos amplicões permitiu esclarecer as dúvidas e concluir que muitas das alterações putativas eram falsos positivos, isto é, que constituíam artefactos das reacções de PCR ou sequenciação.

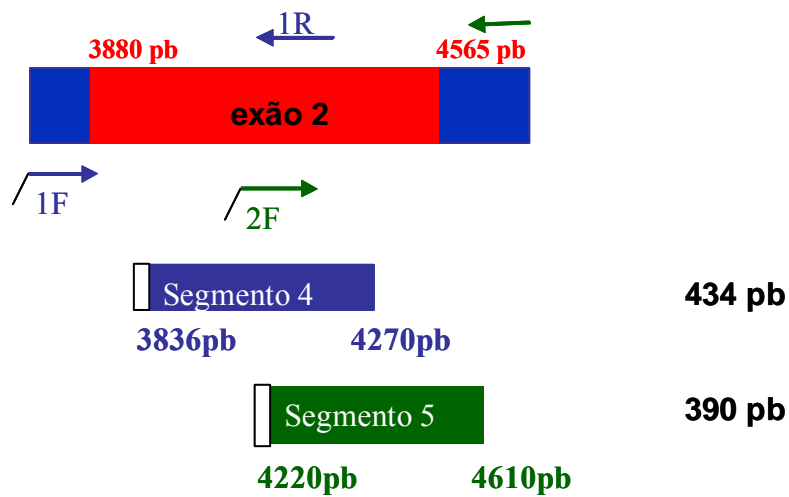


**Figura 34:** Electroforetogramas e representação esquemática das regiões de sobreposição do SIX4. A: comparação entre os segmentos 1-2. B: comparação entre os segmentos 2-3.

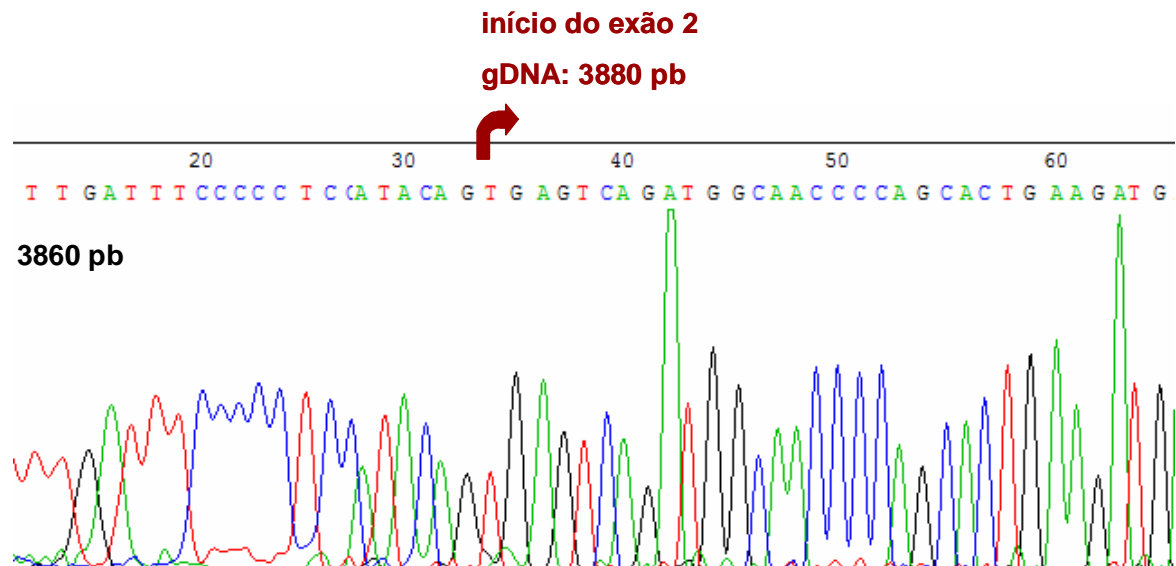
Após a análise, por inspecção visual e com o programa DNASIS v5.0-Demo (Hitachi), e comparação rigorosa entre os diferentes electroforetogramas não foram detectadas alterações somáticas no exão 1, do gene SIX4, nas amostras tumorais estudadas.

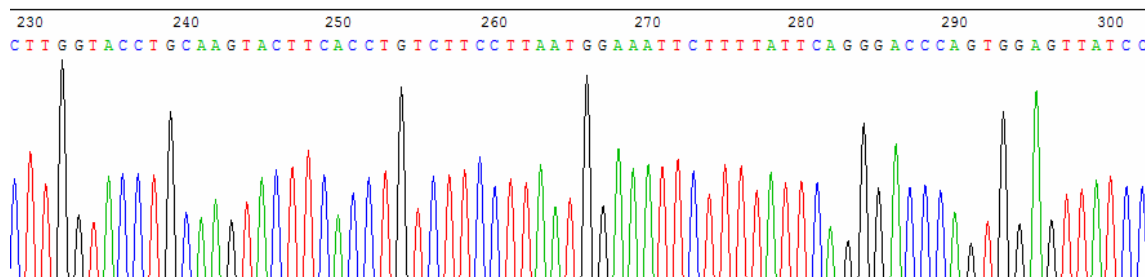
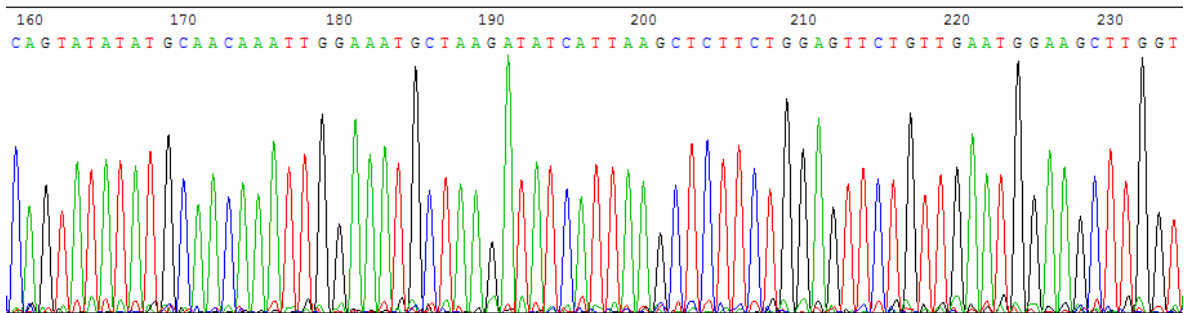
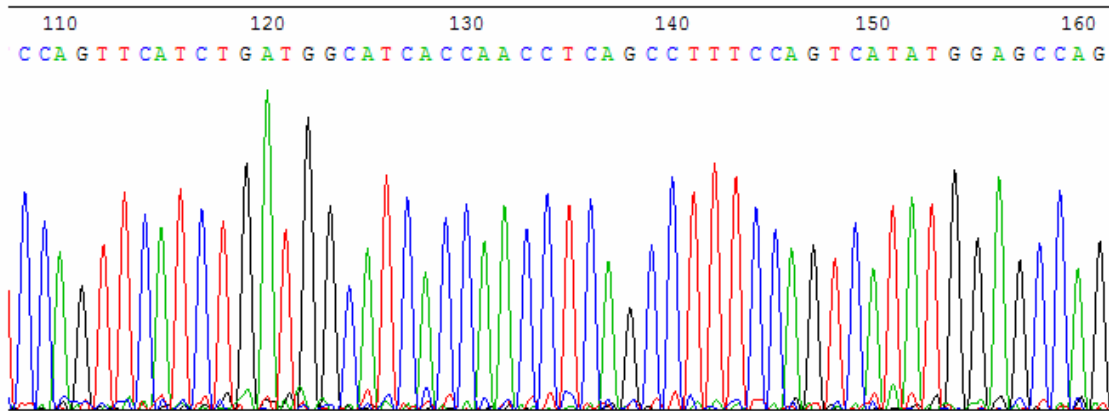
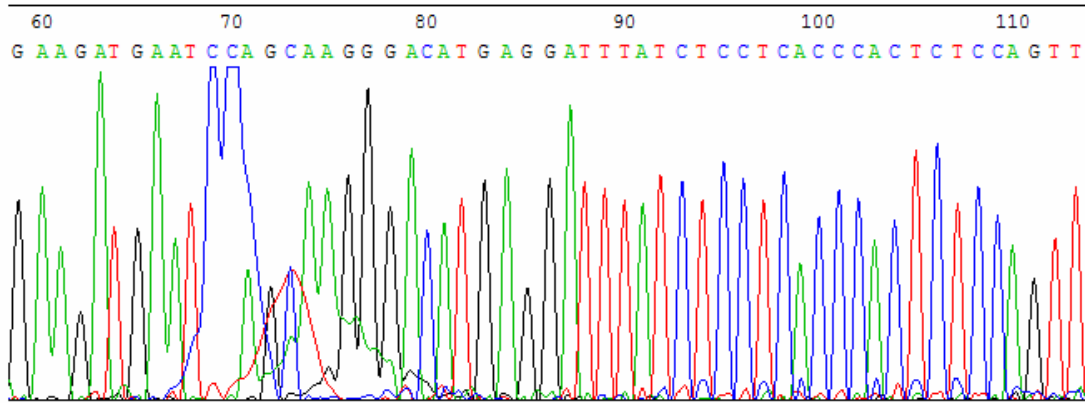
## Pesquisa de alterações somáticas no Exão 2 do SIX4

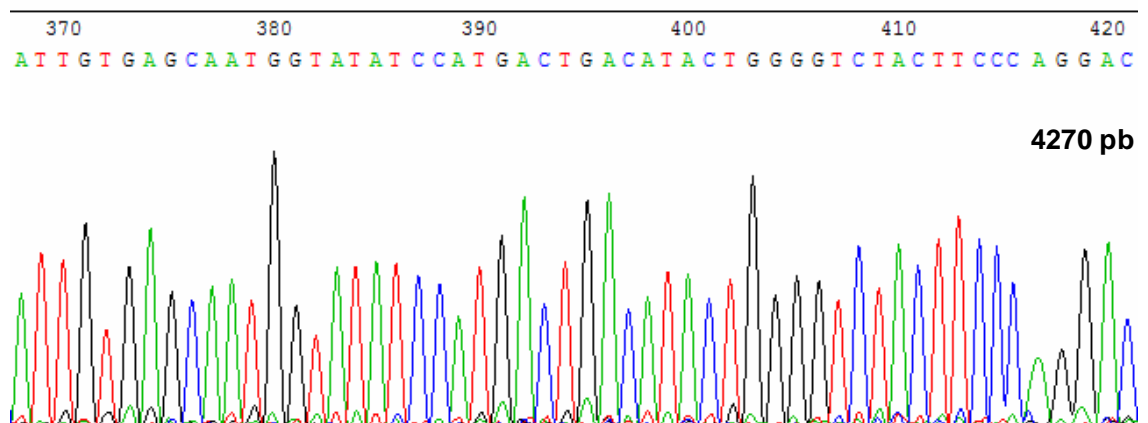
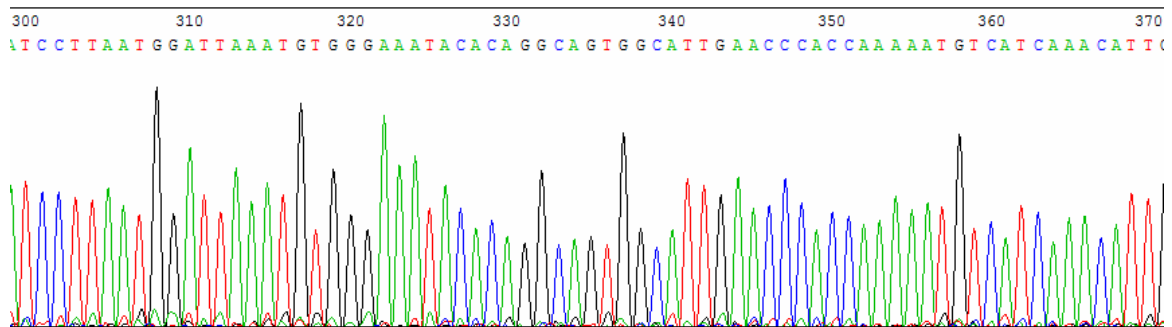
O exão 2 foi amplificado em 2 segmentos com zonas sobrepostas (**Figura 35**), e cada um deles foi sequenciado com o *primer* M13.



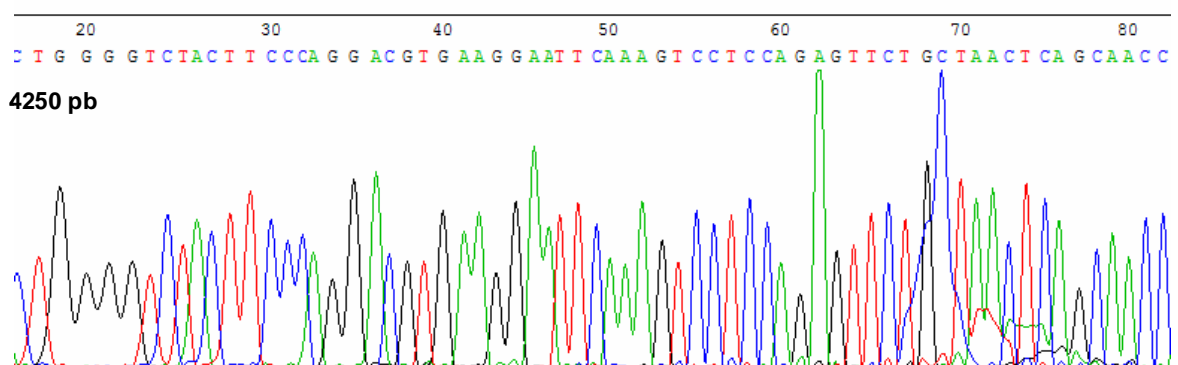
**Figura 35:** Representação esquemática da amplificação do exão 2 do gene SIX4.

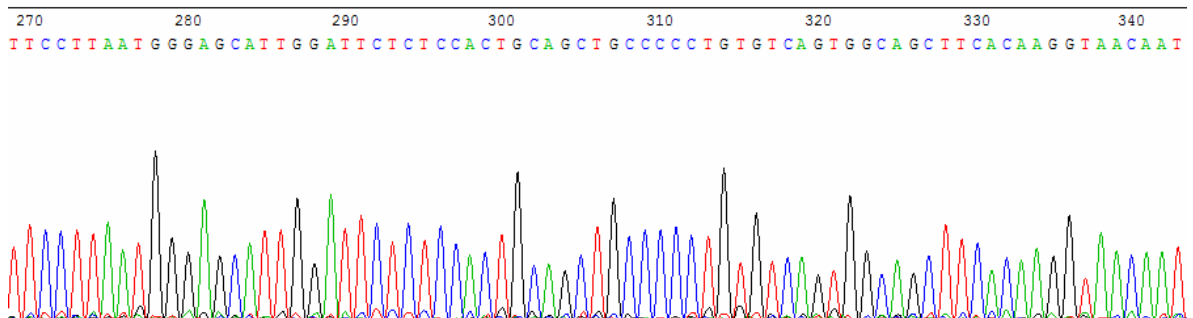
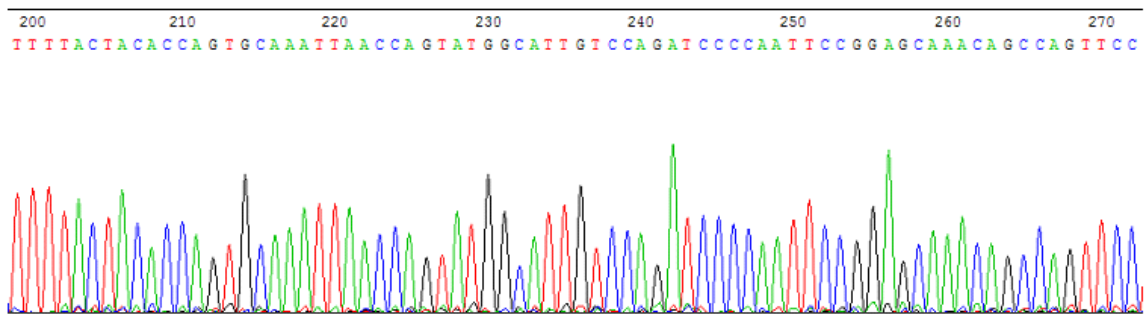
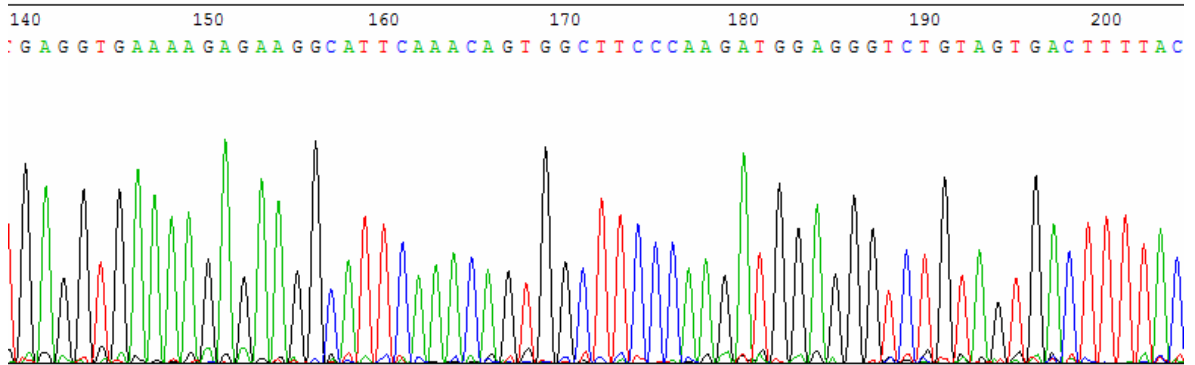
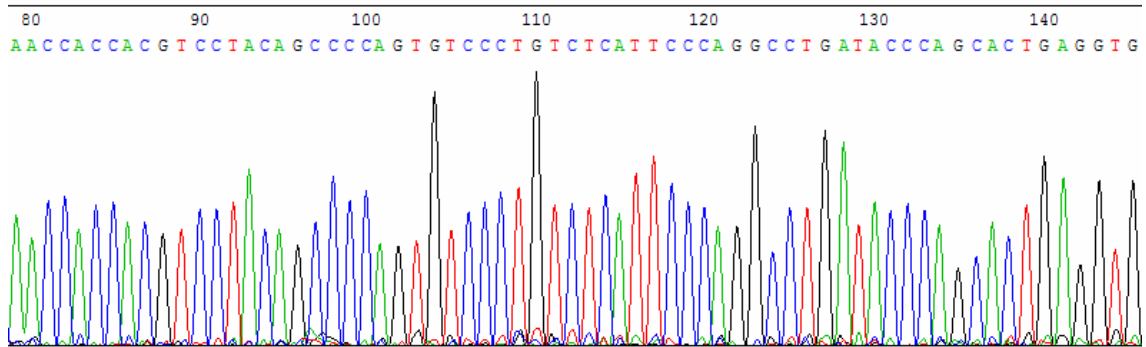


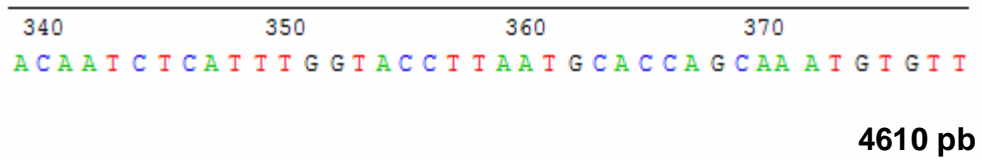




**Figura 36:** Electroforetograma correspondente à sequenciação do amplicão 4, exão 2, do SIX4.





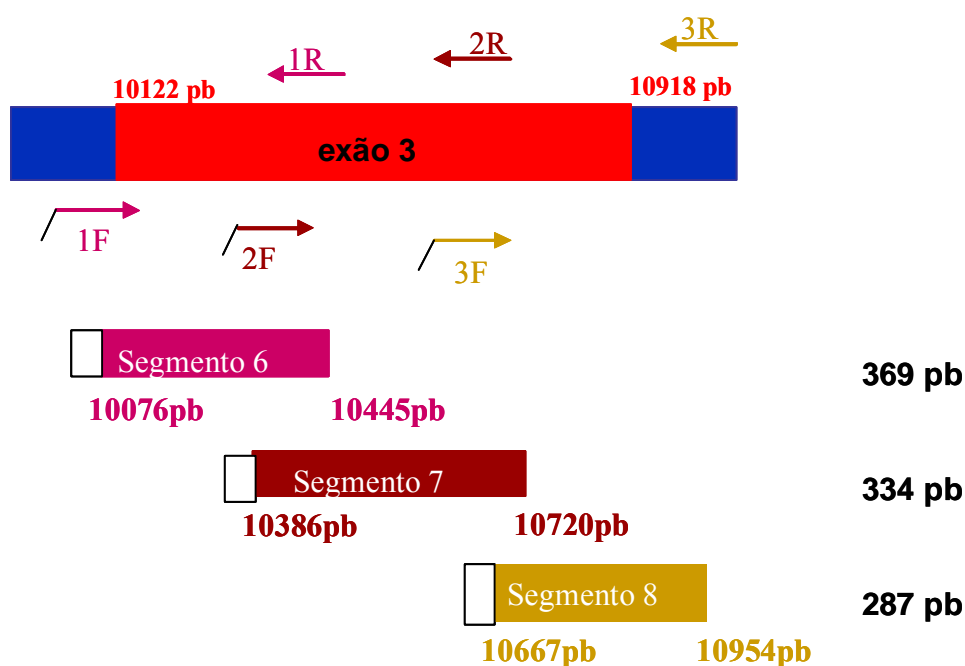


**Figura 37:** Electroforetograma correspondente à sequenciação do amplicão 5, exão 2, do SIX4.

Após a análise, por inspeção visual e com o programa DNASIS v5.0-Demo (Hitachi), e comparação rigorosa entre os diferentes electroforetogramas não foram detectadas alterações somáticas no exão 2, do gene SIX4, nas amostras tumorais estudadas.

### Sequenciação do exão 3 do gene SIX4

O exão 3 foi amplificado em 3 segmentos com zonas sobrepostas (**Figura 38**), e cada um deles foi sequenciado, com o oligómero M13 ou com o oligómero *reverse* específico do amplificação.

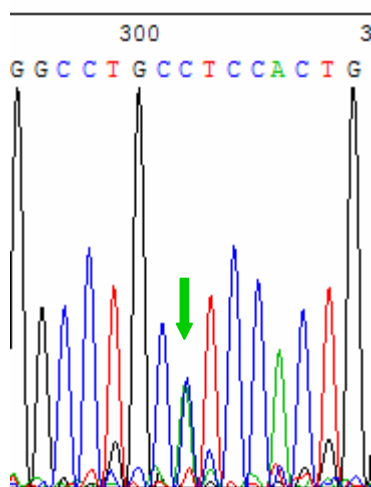


**Figura 38:** Representação esquemática da amplificação do exão 3 do gene SIX4.

Após analisar visualmente os electroforetogramas resultantes da sequenciação do amplificado que cobre o segmento 6, das amostras 1 a 4, observou-se a existência de muito ruído, para as amostras 1 e 2, e verificou-se que não se obtinha um sinal legível para as amostras 3 e 4. Uma vez que os electroforetogramas obtidos dificultavam ou impediam a busca de mutações, resolveu testar-se reacções de sequenciação com o oligómero *reverse* específico do amplificação.

Verificou-se que a qualidade dos electroforetogramas resultantes da sequenciação com o oligómero M13 ou com o *reverse* específico eram idênticas, pelo que o resultado sugere que o DNA é o factor limitante.

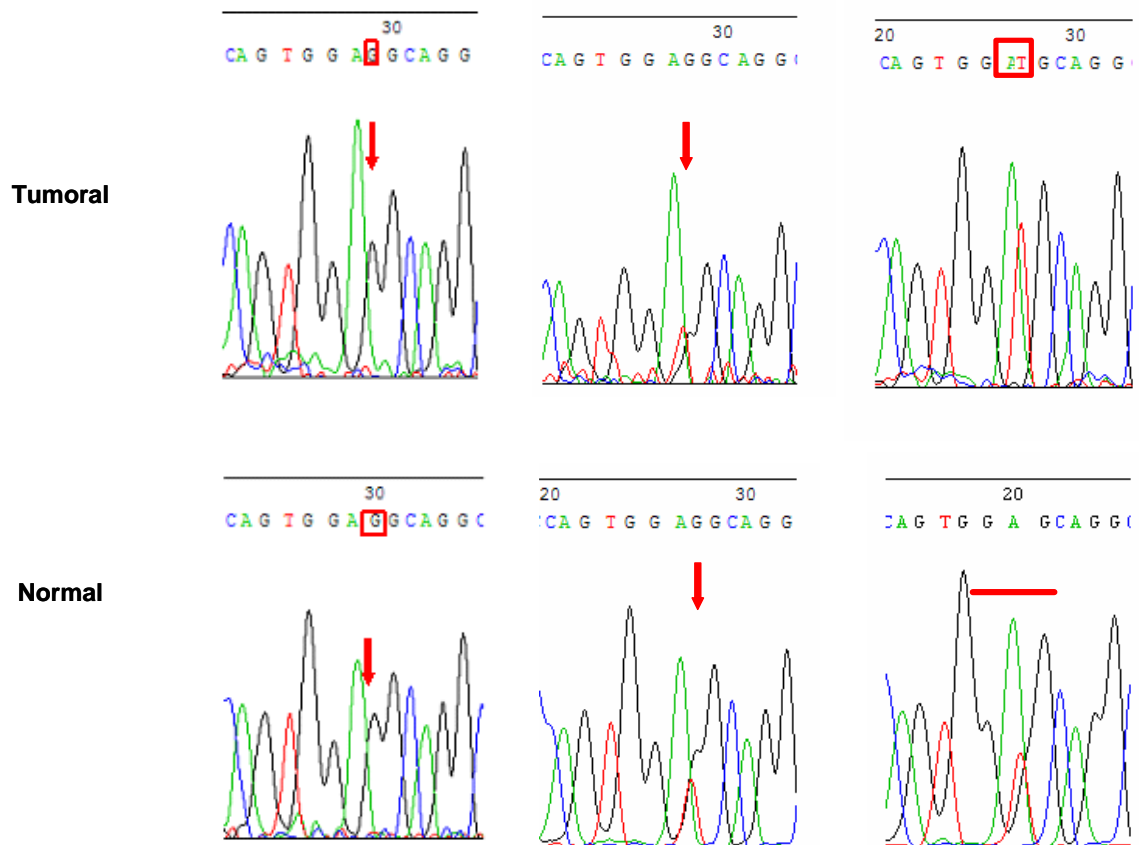
No segmento 6 detectou-se um sinal suspeito de se tratar de uma alteração somática (**Figura 39**). Para confirmar a presença de alteração somática, sequenciou-se o gDNA extraído do tecido normal circundante ao tecido tumoral, do mesmo paciente.



**Figura 39:** Electroforetograma onde é visível um sinal que sugere a presença de uma alteração somática.

Verificou-se que o sinal sugestivo de presença de alteração somática também estava presente no electroforetograma resultante da sequenciação do gDNA extraído do tecido normal –controlo– (**Figura 40**), pelo que se concluiu que esta sequência é polimórfica e que se tratava de uma variante hereditária.

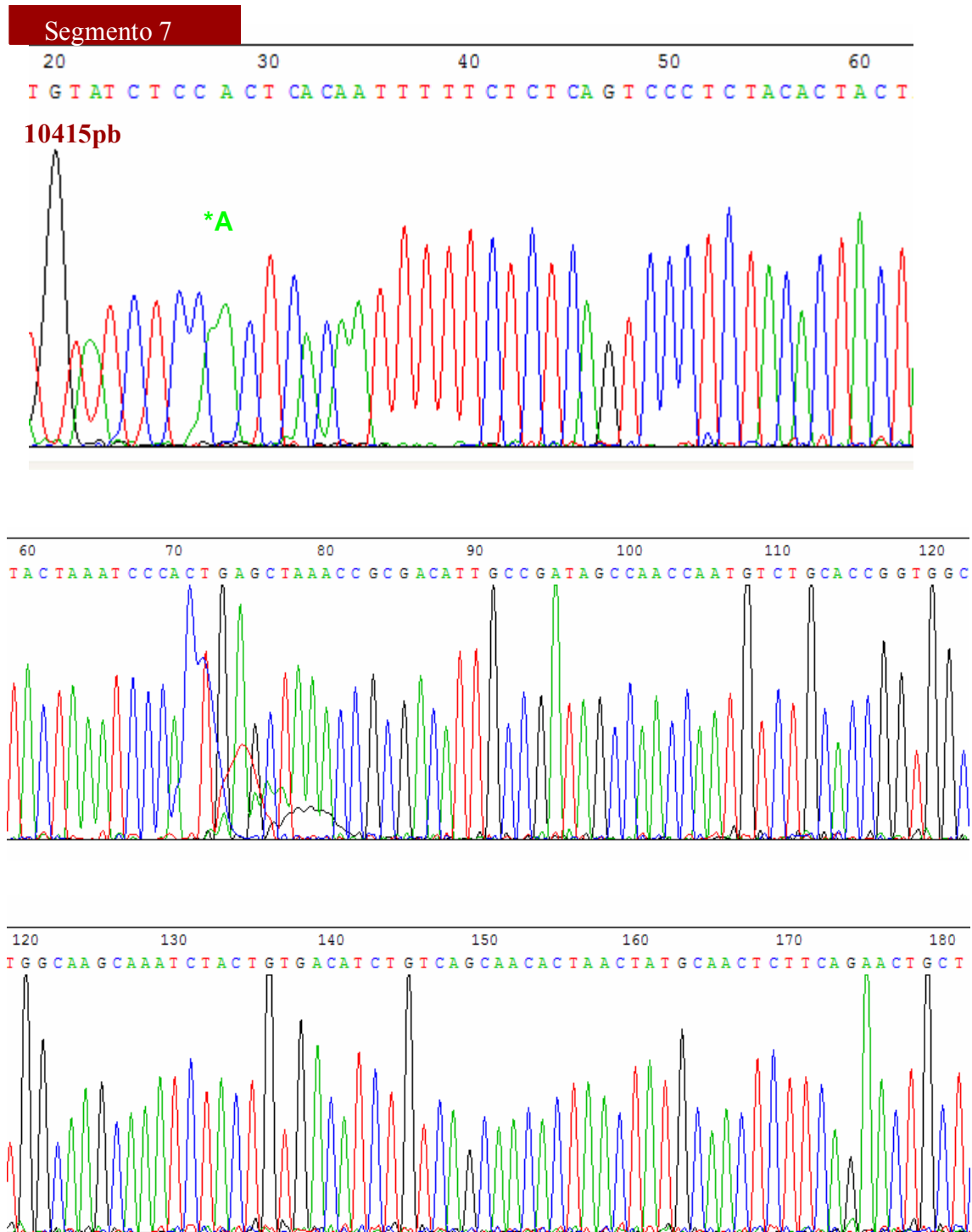
A frequência alélica (**Tabela 3**) indica que esta variante é comum na população de pacientes com carcinoma ductal da mama analisados. Embora seja apelativo especular sobre a potencial susceptibilidade para o carcinoma da mama causada por este alelo, a base de dados, [http://www.hapmap.org/cgiiperl/gbrowse/hapmap\\_B35/](http://www.hapmap.org/cgiiperl/gbrowse/hapmap_B35/), sobre genótipos de populações e frequências alélicas, mostra que esta variante já foi identificada e que a frequência dos alelos é variável entre populações.

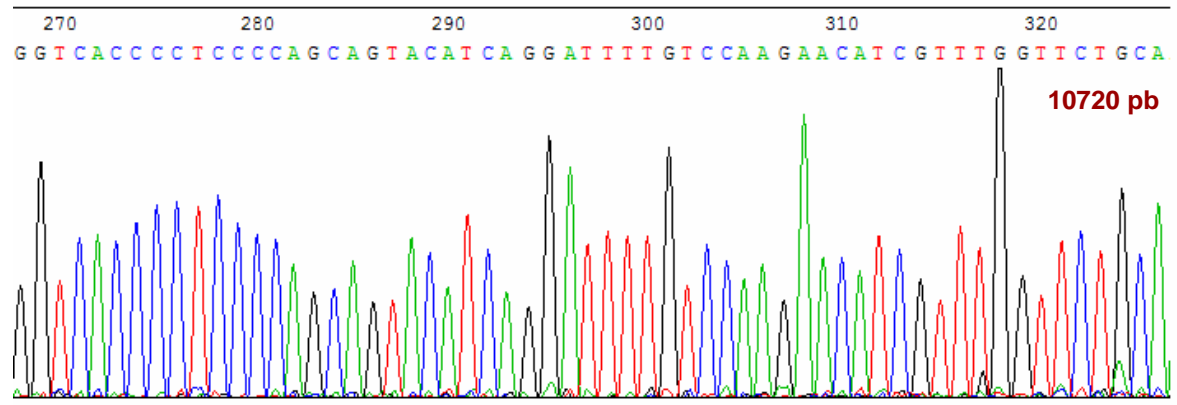
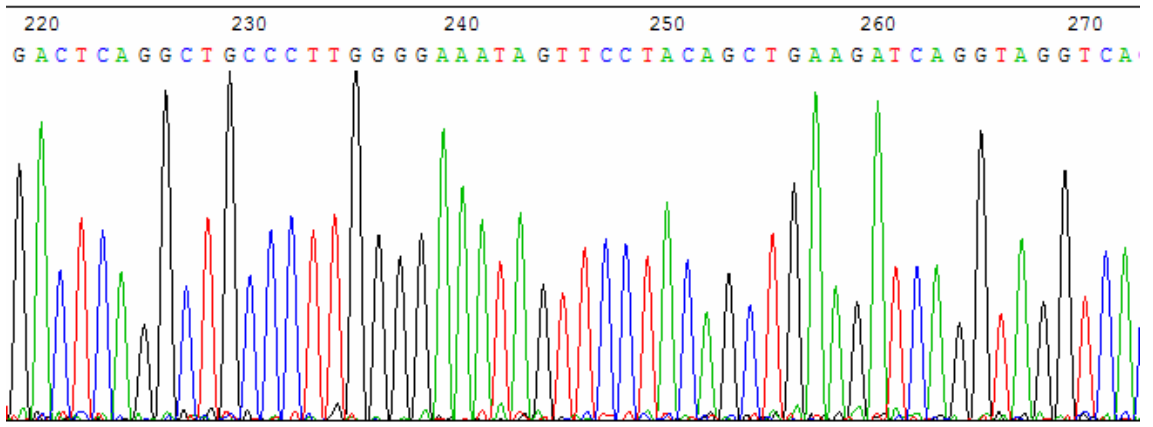
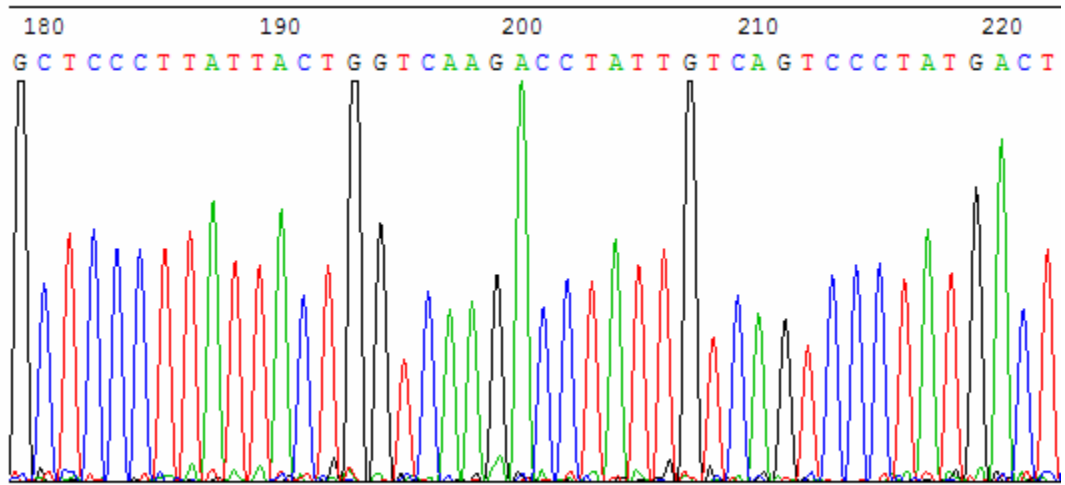


**Figura 40:** Electroforetogramas do amplicão 6 – exão 3, SIX4 – que confirmam a presença de uma alteração, em relação à sequência de referência.

Número de casos	Alelo 1	Alelo 2	%
13	C	A	38
14	C	C	41
7	A	A	21

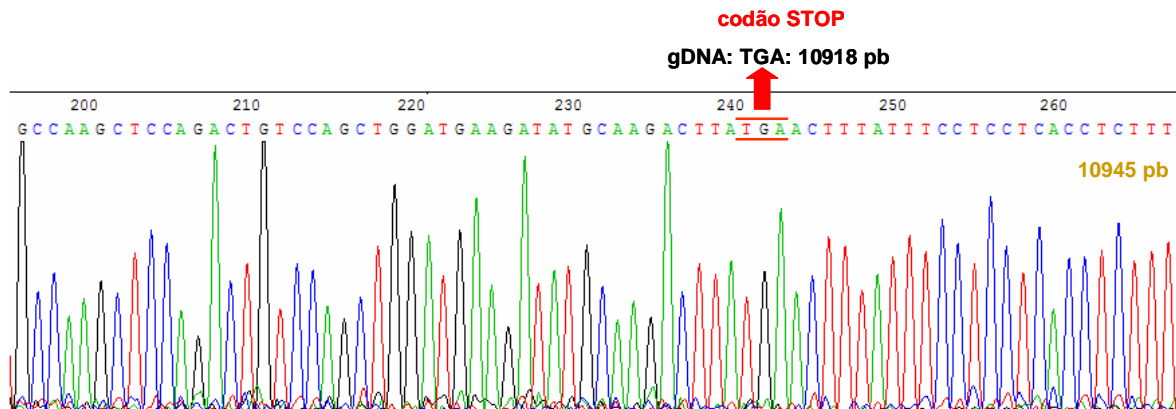
**Tabela 3:** Frequência alélica no grupo de amostras de carcinomas da mama analisados.





**Figura 41:** Electroforetograma correspondente à sequenciação do amplicão 7 –exão 3, SIX4.





**Figura 42:** Electroforetograma correspondente à sequenciação do amplicão 8, exão 3, do gene SIX4.

Após a análise, por inspecção visual e com o programa DNASIS v5.0-Demo (Hitachi), e comparação rigorosa entre os diferentes electroforetogramas não foram detectadas alterações somáticas no exão 3, do gene SIX4, nas amostras tumorais estudadas.

Dos 48 casos sequenciados obtiveram-se 40, 34, 48, 44, 37, 34, 46 e 44 electroforetogramas, respectivamente para os amplicões 1 a 8, com qualidade para se pesquisarem alterações somáticas. Verificámos que a estratégia escolhida, de utilizar um *primer universal* é adequada para a sequenciação de um grande número de amostras, embora por vezes, provavelmente devido a características inerentes à própria sequência nucleotídica ou à qualidade do DNA, seja recomendado usar o oligómero reverse.

O objectivo proposto de busca de alterações somáticas no gene SIX4 foi atingido, concluindo-se que não há mutações somáticas no gene SIX4, nas amostras tumorais analisadas. Com base no resultado obtido pode especular-se que as mutações no SIX4 não são frequentes, mas não se deve excluir a hipótese do gene estar envolvido na tumorigénese da mama. Um dos exemplos que ilustra o que atrás foi dito é o CDX2. Embora raramente sejam encontradas mutações no CDX2 (da Costa LT *et al*, 1999; Sivagnanasundaram *et al*, 2001; Wicking *et al*, 1998), a perda da sua função causa predisposição para o carcinoma colo-rectal (Ee *et al*, 1995). Quer o SIX4 quer o CDX2

são genes homeobox e raramente se encontram mutações em tumores nesta família de genes (Goodman & Scambler, 2001).

Nas amostras analisadas detectou-se uma alteração não-sinónima e embora esta variante alélica já tenha sido descrita, não se pode descartar a hipótese desta sequência polimórfica estar associada a susceptibilidade para carcinoma da mama. No caso do CDX2 e do carcinoma colo-rectal, é muito frequente observar-se a ausência de expressão da proteína. Uma vez que foi detectada uma alteração não-sinónima no exão 3, que é onde se localiza o domínio de activação de transcrição do SIX4, é importante testar a capacidade de transactivação e os níveis de expressão da proteína, quer para o SIX4 com a sequência de referência quer para o SIX4 com a variante hereditária e com as alterações somáticas anteriormente encontradas em carcinoma da mama humano (Sjöblom T, *et al*, 2006).

**Investigação em curso**  
**&**  
**Perspectivas futuras**

## Estudos funcionais com o SIX4

O SIX4 é um gene de regulação de transcrição em que foram detectadas mutações no carcinoma da mama (Sjöblom T *et al*, 2006). Embora nas amostras analisadas não tenham sido detectadas alterações somáticas, foi detectada uma alteração não-sinónima no exão 3, o exão que codifica o domínio de activação de transcrição do SIX4. Para esclarecer o papel do SIX4 no carcinoma da mama, é relevante avaliar as consequências funcionais desta variante alélica e também das mutações somáticas anteriormente detectadas (Sjöblom T *et al*, 2006).

Com a investigação em curso, pretendemos avaliar a capacidade de transactivação do SIX4 e os níveis de expressão da proteína. Para atingir o objectivo proposto, planeámos uma estratégia que envolve a construção de um vector repórter análogo aos existentes para os TCFs (da Costa LT *et al*, 1999), em que o cDNA da Luciferase de *Photinus pyralis* irá ficar sob o controlo de um promotor contendo quatro repetições concatamerizadas da sequência de ligação do SIX4 ao DNA e de um vector de activação, no qual o cDNA de SIX4 fica sob o controlo de um promotor constitutivo em células de mamífero. Esses vectores serão utilizados em ensaios de actividade de luciferase em culturas celulares humanas co-transfectadas em simultâneo com três plasmídeos distintos: um repórter da eficiência de transfecção (CMV $\beta$ -Gal), um repórter da actividade do SIX4 (4Six4BS-Luc) e um de vários vectores de expressão de SIX4 – com o SIX4 normal, a variante hereditária ou cada uma das mutações somáticas anteriormente detectadas (Sjöblom T *et al*, 2006). Os vectores de expressão das diferentes formas do SIX4 serão igualmente utilizados para avaliar a influência das diferentes variantes (hereditária e somáticas) do gene no nível de expressão da proteína.

Foram desenhados oligómeros com homologia à sequência nucleotídica reconhecida pelo domínio de ligação ao DNA de proteínas da família SIX, de modo a gerarmos um concatâmero constituído por quatro repetições.

```

DNA sequence bound by Six2/Six4/Six5
TCGAGCCGGTGTTCAGGTTGCTCC
    
```

(Kawakami K *et al*, 1996 b)

Resumidamente, desenharam-se oligómeros complementares às extremidades da sequência que se pretendia *concatamerizar*, e complementares entre si, cada um dos quais com uma “cauda” *BamHI* ou *BglIII*.

Foi preparada uma mistura contendo cada um dos oligómeros, que foi colocada num banho térmico a 50°C, durante 15 minutos, de modo às cadeias hibridizarem.

Após o arrefecimento lento, até aproximadamente os 12 °C, foram adicionados dNTPs e T4 DNA polimerase (MBI Fermentas, Vilnius, Lituânia), formando-se uma cadeia dupla.

07-89: 1F: gat AGA TCT CGA G **CC GGT GTC AGG TTG C**  
28nn

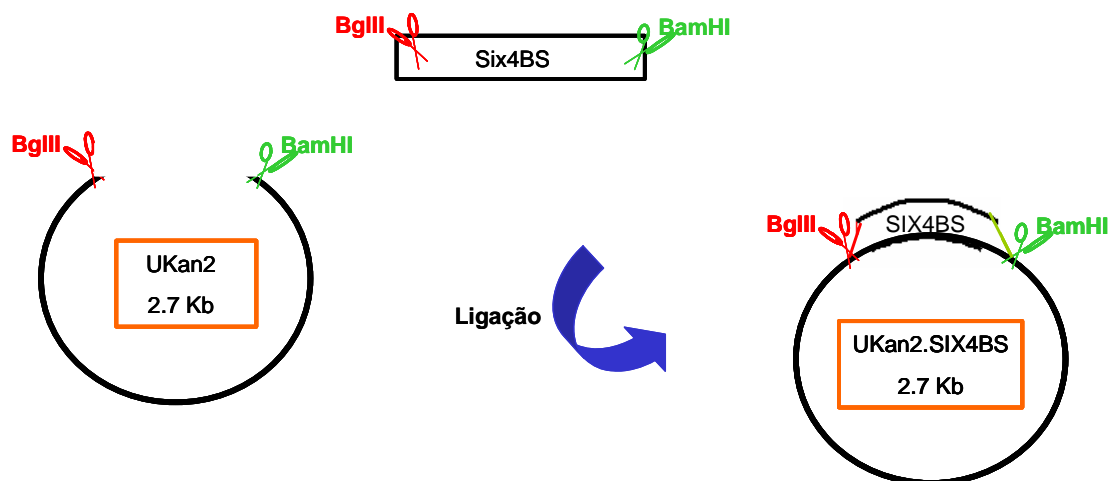


07-90: 1R: GTT **GGA TCC GGA** **GCA ACC TGA CAC CGG**  
27nn



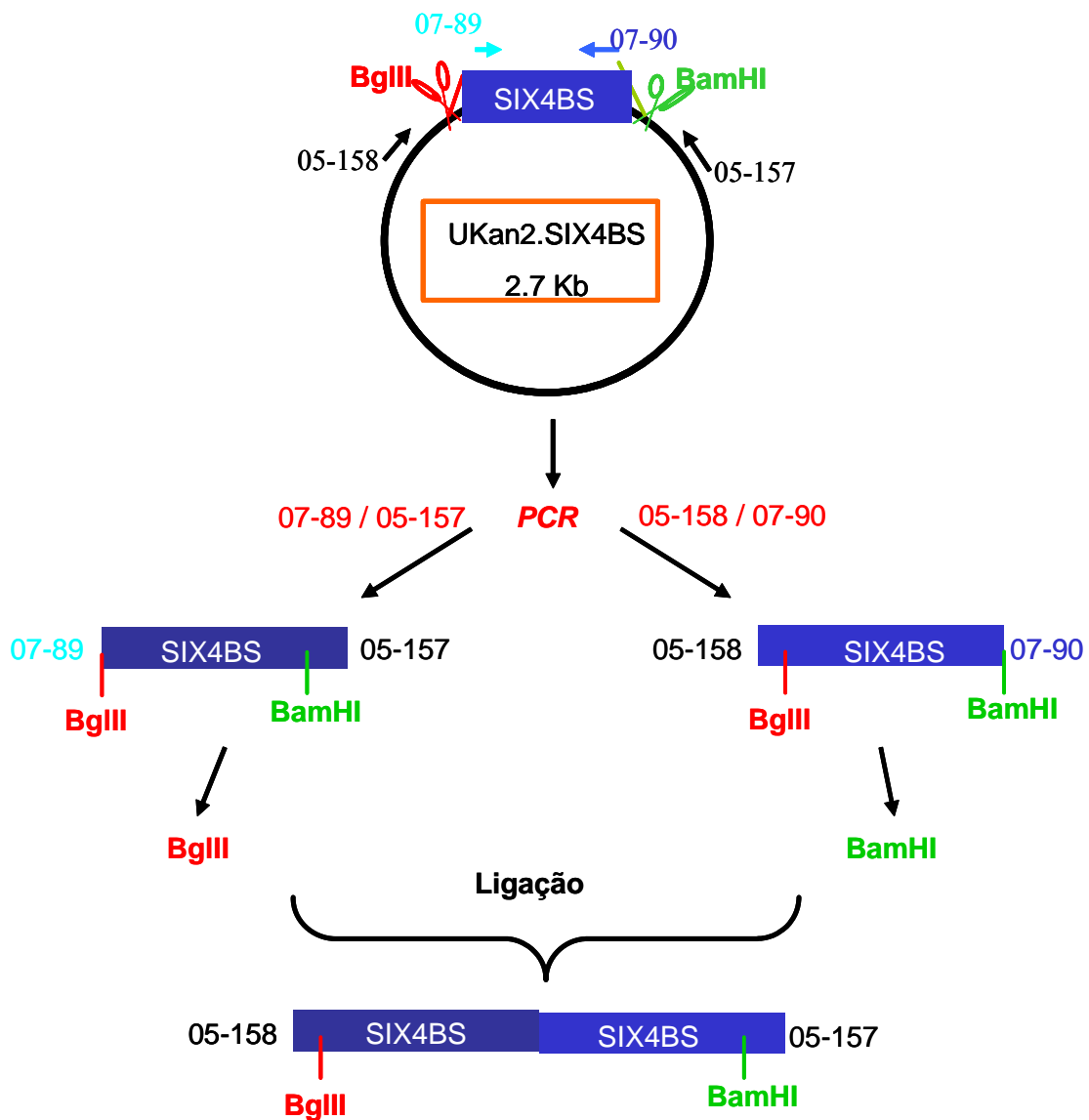
**Figura 43:** Representação esquemática da formação da sequência nucleotídica reconhecida pelo domínio de ligação ao DNA de proteínas da família SIX por hibridização dos oligômeros e síntese pela T4DNA polimerase.

A cadeia dupla assim gerada foi digerida com as enzimas de restrição *BamHI* ou *BglIII* e clonada no vector UKan2-TA, digerido com as mesmas (**Figura 44**).



**Figura 44:** Representação esquemática da clonagem da sequência nucleotídica reconhecida pelo domínio de ligação ao DNA do SIX4 (SIX4 Binding Site: Six4BS) no vector UKan.2TA *BamHI*/ *BglIII*.

A sequência do fragmento Six4BS clonado em UKan2-TA foi confirmada por sequenciação e o UKan2.SIX4BS utilizado como substrato em duas reações de *PCR* distintas. Cada um dos produtos amplificados foi digerido com as enzimas de restrição *BglIII* ou *BamHI*. Assim, geraram-se fragmentos com extremidades compatíveis e cada um com um local de restrição para *BglIII* ou *BamHI* intacto (**Figura 45**).



**Figura 45:** Representação esquemática da formação de um concâmero com 2 cópias do Six4BS.

Estes foram ligados e o produto de ligação amplificado por *PCR*, com o par de oligômeros 05-157/ 05-158. O produto resultante foi digerido com *BamHI* e *BglIII*, clonado no vector UKan2-TA (e a sua sequência confirmada) dando origem a UKan2.2Six4BS.

O UKan2.2Six4BS foi usado como molde numa reacção de *PCR* com o par de oligómeros 05-157/ 05-158 e, em reacções paralelas, o produto resultante foi digerido com *Bam*HI ou *Bgl*II e os fragmentos resultantes ligados (Figura 46).

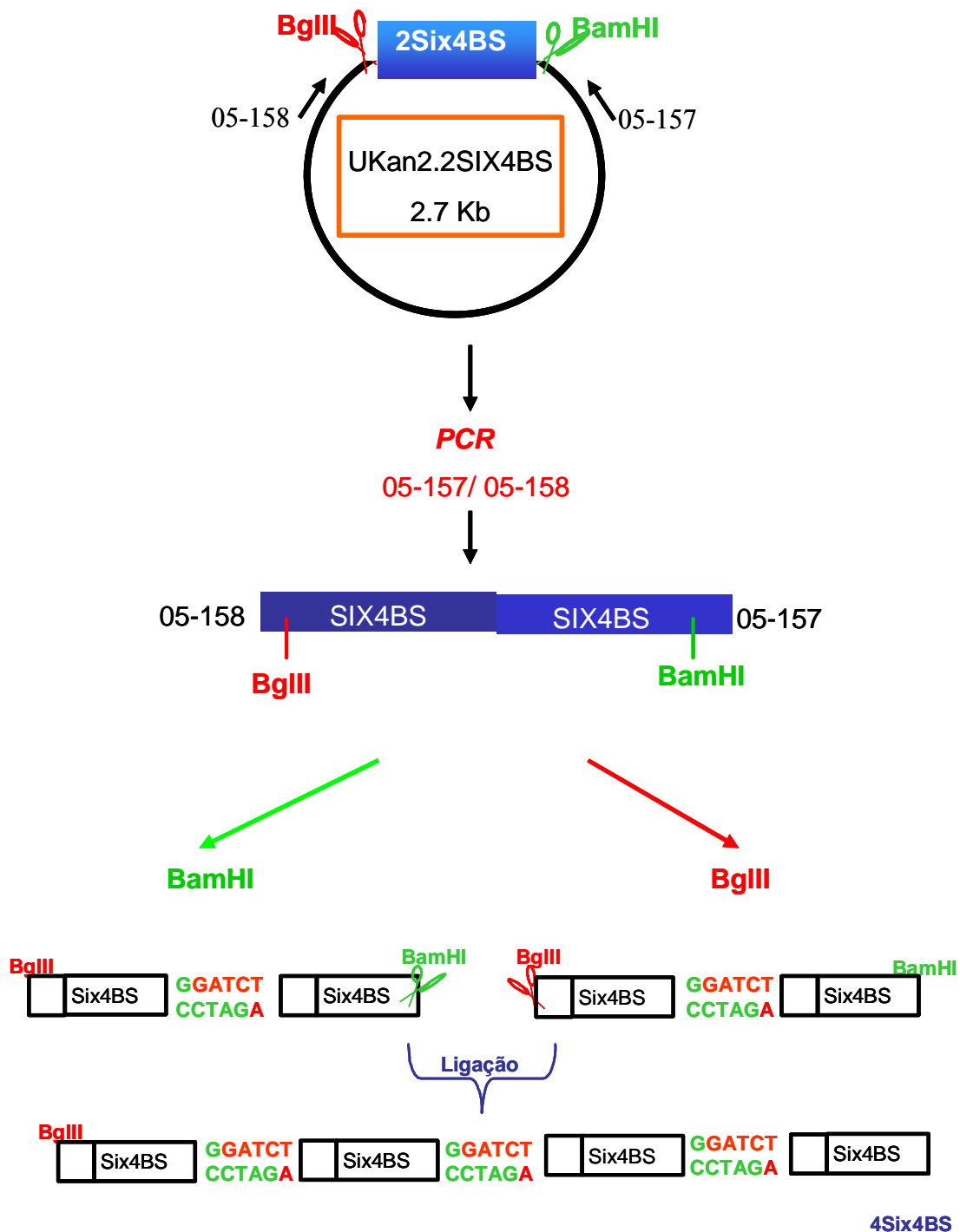


Figura 46: Representação esquemática da formação do concânamero 4 Six4BS.

O fragmento com 4 cópias da sequência nucleotídica reconhecida pelo domínio de ligação ao DNA do SIX4 – o 4 Six4BS – foi seguidamente amplificado por *PCR*, com o par de oligómeros 05-157/05-158 e clonado em UKan2-TA *Bam*HI/ *Bgl*III do modo análogo aos fragmentos anteriores, gerando-se assim o UKan2.4Six4BS (**Figura 47**), e posteriormente usado para substituir as três cópias da sequência reconhecida pelos TCFs em OT-Luc por quatro cópias da sequência de ligação ao DNA de Six4, gerando-se assim o repórter 4Six4BS-Luc.

Para a construção dos vectores de expressão do Six4, toda a zona codificante do gene foi amplificada em 5 segmentos de aproximadamente 500pb cada, usando os oligómeros indicados na **Tabela II, em anexo**. Estes serão clonados em UKAn2-TA e, após confirmação das respectivas sequências, reunidos por digestão com enzimas apropriadas e ligação. A zona codificante completa do SIX4 assim obtida será então transferida para o vector de expressão CMV-Myc (Clontech, Palo Alto, EUA), gerando o vector de expressão do Six4 normal (CMV-Myc.Six4). Os vectores de expressão das restantes variantes do Six4 serão posteriormente gerados a partir de CMV-Myc.Six4 por mutagénesis dirigida.

## Bibliografia

Abate-Shen C (2002) “Deregulated homeobox gene expression in cancer: cause or consequence?” *Nat Rev Cancer* 2 (10): 777-85.

Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF (2002) “A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes.” *Br J Cancer*; 86 (1): 76-83.

Asselin-Labat ML, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, Hartley L, Robb L, Grosveld FG, van der Wees J, Lindeman GJ, Visvader JE (2007) “Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation.” *Nat Cell Biol.* 9 (2): 201-9.

Barsky SH & Karlin NJ (2005) “Myoepithelial cells: autocrine and paracrine suppressors of breast cancer progression.” *J Mammary Gland Biol Neoplasia*; 10 (3): 249-60.

Berns EM, Foekens JA, Vossen R, Look MP, Devilee P, Henzen-Logmans SC, van Staveren IL, van Putten WL, Inganäs M, Meijer-van Gelder ME, Cornelisse C, Claassen CJ, Portengen H, Bakker B, Klijn JG (2000) “Complete sequencing of TP53 predicts poor response to systemic therapy of advanced breast cancer.” *Cancer Res.* 60 (8): 2155-62.

Böcker W, Moll R, Poremba C, Holland R, Van Diest PJ, Dervan P, Bürger H, Wai D, Ina Diallo R, Brandt B, Herbst H, Schmidt A, Lerch MM, Buchwallow IB (2002) “Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: a new cell biological concept.” *Lab Invest.*; 82 (6): 737-46.

Brenner AJ & Aldaz CM (1997) "The genetics of sporadic breast cancer." *Prog Clin Biol Res.*; 396: 63-82.

Chen H & Sukumar S (2003) "Role of homeobox genes in normal mammary gland development and breast tumorigenesis." *J Mammary Gland Biol Neoplasia*; 8 (2): 159-75.

Cheyette BN, Green PJ, Martin K, Garren H, Hartenstein V, Zipursky SL (1994) "The *Drosophila sine oculis* locus encodes a homeodomain-containing protein required for the development of the entire visual system." *Neuron*. 12 (5): 977-96.

Cho KR & Vogelstein B (1992) "Genetic alterations in the adenoma--carcinoma sequence." *Cancer* 70 (6 Suppl): 1727-31.

da Costa LT, He TC, Yu J, Sparks AB, Morin PJ, Polyak K, Laken S, Vogelstein B, Kinzler KW (1999) "CDX2 is mutated in a colorectal cancer with normal APC/beta-catenin signaling." *Oncogene*; 18 (35): 5010-4.

da Costa LF (2001) "Return of de-differentiation: why cancer is a developmental disease" *Curr Opin Oncol*. 13 (1): 58-62.

Dahiya R & Deng G (1998) "Molecular prognostic markers in breast cancer." *Breast Cancer Res Treat.*; 52 (1-3): 185-200.

Ee HC, Erler T, Bhathal PS, Young GP, James RJ ( 1995) "Cdx-2 homeodomain protein expression in human and rat colorectal adenoma and carcinoma." *Am J Pathol.*; 147 (3): 586-92.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) "Cluster analysis and display of genome-wide expression patterns." *Proc Natl Acad Sci U S A*. 95 (25): 14863-8.

Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA (1996) "Laser capture microdissection." *Science*; 274 (5289): 998-1001.

Ferlay J, Autier P, Boniol M, Heanue M, Colombet M, Boyle P (2007) "Estimates of the cancer incidence and mortality in Europe in 2006." *Ann Oncol.* 18 (3): 581-92.

Ford HL, Landesman-Bollag E, Dacwag CS, Stukenberg PT, Pardee AB, Seldin DC. (2000) "Cell cycle-regulated phosphorylation of the human SIX1 homeodomain protein." *J Biol Chem.*; 275 (29): 22245-54.

Ford HL, Kabingu EN, Bump EA, Mutter GL, Pardee AB (1998) "Abrogation of the G2 cell cycle checkpoint associated with overexpression of HSIX1: a possible mechanism of breast carcinogenesis." *Proc Natl Acad Sci U S A.*; 95 (21): 12608-13.

Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-Perez D, Affolter M, Otting G, Wüthrich K "Homeodomain-DNA recognition." (1994) *Cell*; 78 (2): 211-23.

Goldberg DM & Diamandis EP (1993) "Models of neoplasia and their diagnostic implications: a historical perspective." *Clin Chem*; 39 (11 Pt 2): 2360-74.

Hanahan D & Weinberg RA (2000) "The hallmarks of cancer" *Cell* 100 (1): 57-70.

Kasahara H, Izumo S (1999) "Identification of the in vivo casein kinase II phosphorylation site within the homeodomain of the cardiac tissue-specifying homeobox gene product Csx/Nkx2.5." *Mol Cell Biol.* 19 (1): 526-36.

Kawakami K, Sato S, Ozaki H, Ikeda K. (2000) "Six family genes--structure and function as transcription factors and their roles in development." *Bioessays*; 22 (7): 616-26.

Kawakami K, Ohto H, Ikeda K, Roeder RG (1996 a) "Structure, function and expression of a murine homeobox protein AREC3, a homologue of *Drosophila sine oculis* gene product, and implication in development." *Nucleic Acids Res.* 24 (2): 303-10.

Kawakami K, Ohto H, Takizawa T, Saito T (1996 b) "Identification and expression of six family genes in mouse retina." *FEBS Lett.*; 393 (2-3): 259-63.

Knudson AG (2002) "Cancer genetics." *Am J Med Genet.* 111 (1): 96-102.

Kobayashi H, Kawakami K, Asashima M, Nishinakamura R (2007) "Six1 and Six4 are essential for Gdnf expression in the metanephric mesenchyme and ureteric bud formation, while Six1 deficiency alone causes mesonephric-tubule defects." *Mech Dev.*;124(4):290-303.

Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z (2006) "GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland." *Cell.* 127 (5): 1041-55.

Lakhani SR, O'Hare MJ (2001) "The mammary myoepithelial cell--Cinderella or ugly sister?" *Breast Cancer Res.*; 3(1): 1-4.

Laughon A (1991) "DNA binding specificity of homeodomains." *Biochemistry*; 30 (48): 11357-67.

Lester SC (2005) "The Breast" in Robbins & Cotran Pathologic Basis of Disease; 7 th edition; Pennsylvania, Elsevier Inc.; pp249 a 254.

Lewis MT (2000) "Homeobox genes in mammary gland development and neoplasia." *Breast Cancer Res.* 2 (3): 158-69.

McGinnis W, Hart CP, Gehring WJ, Ruddle FH (1984) "Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of *Drosophila*." *Cell*; 38 (3): 675-80.

McGinnis W, Krumlauf R (1992) "Homeobox genes and axial patterning." *Cell.* ; 68 (2): 283-302.

Nowell PC (1976) "The clonal evolution of tumor cell populations." *Science* 194 (4260): 23-8.

Ohto H, Kamada S, Tago K, Tominaga SI, Ozaki H, Sato S, Kawakami K (1999) "Cooperation of six and eya in activation of their target genes through nuclear translocation of Eya." *Mol Cell Biol.*; 19 (10): 6815-24.

Ozaki H, Watanabe Y, Takahashi K, Kitamura K, Tanaka A, Urase K, Momoi T, Sudo K, Sakagami J, Asano M, Iwakura Y, Kawakami K (2001) "Six4, a putative myogenin gene regulator, is not essential for mouse embryonal development." *Mol Cell Biol.*; 21 (10): 3343-50.

Ozaki H, Yamada K, Kobayashi M, Asakawa S, Minoshima S, Shimizu N, Kajitani M, Kawakami K. (1999) "Structure and chromosome mapping of the human SIX4 and murine Six4 genes." *Cytogenet Cell Genet.* 87 (1-2): 108-12.

Parkin DM, Bray F, Ferlay J, Pisani P (2005) "Global cancer statistics, 2002." *CA Cancer J Clin.* 55 (2): 74-108.

Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D. (2000) "Molecular portraits of human breast tumours." *Nature* 406 (6797): 747-52.

Peto J. (2002) "Breast cancer susceptibility-A new look at an old model." *Cancer Cell.*; 1 (5): 411-2.

Pignoni F, Hu B, Zavitz KH, Xiao J, Garrity PA, Zipursky SL (1997) "The eye-specification proteins So and Eya form a complex and regulate multiple steps in Drosophila eye development." *Cell*; 91 (7): 881-91.

Ross JS, Fletcher JA, Linette GP, Stec J, Clark E, Ayers M, Symmans WF, Puzstai L, Bloom KJ (2003) "The Her-2/neu gene and protein in breast cancer 2003: biomarker and target of therapy." *Oncologist*; 8 (4): 307-25.

Santarosa M & Ashworth A (2004) "Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way." *Biochim Biophys Acta*. 1654 (2): 105-22.

Schmitt F & Gobbi H (2006) "Mama" in Filho GB (ed) *Bogliolo Patologia*; 7ed, Belo Horizonte; pp 613-643.

Seo HC, Curtiss J, Mlodzik M, Fjose A (1999) "Six class homeobox genes in drosophila belong to three distinct families and are involved in head development." *Mech Dev.*; 83 (1-2): 127-39.

Serikaku MA & O'Tousa JE (1994) "*sine oculis* is a Homeobox Gene Required for Drosophila Visual System Development" *Genetics* 138 (4): 1137-50.

Sivagnanasundaram S, Islam I, Talbot I, Drummond F, Walters JR, Edwards YH (2001) "The homeobox gene CDX2 in colorectal carcinoma: a genetic analysis." *Br J Cancer.*; 84 (2): 218-25.

Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J,

Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE (2006) "The consensus coding sequences of human breast and colorectal cancers." *Science*; 314 (5797): 268-74.

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL (1987) "Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene." *Science*; 235 (4785): 177-82.

Sørbye T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale AL. (2001) "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." *Proc Natl Acad Sci U S A.*, 98 (19): 10869-74.

Struhl G (1981) A homoeotic mutation transforming leg to antenna in *Drosophila*. *Nature*; 292 (5824): 635-8.

Sweadner KJ (1989) "Isozymes of the Na<sup>+</sup>/K<sup>+</sup>-ATPase.>"; *Biochim Biophys Acta.* 988(2):185-220.

Thull DL & Vogel VG (2004) "Recognition and management of hereditary breast cancer syndromes." *Oncologist*; 9(1):13-24.

Tlsty TD, Hein PW (2001) "Know thy neighbor: stromal cells can contribute oncogenic signals." *Curr Opin Genet Dev.* 11(1): 54-9.

Tong Q & Hotamisligil GS (2007) "Developmental biology: cell fate in the mammary gland." *Nature*; 445 (7129): 724-6.

Tonin P, Weber B, Offit K, Couch F, Rebbeck TR, Neuhausen S, Godwin AK, Daly M, Wagner-Costalos J, Berman D, Grana G, Fox E, Kane MF, Kolodner RD, Krainer M, Haber DA, Struwing JP, Warner E, Rosen B, Lerman C, Peshkin B, Norton L, Serova O, Foulkes WD, Garber JE, *et al.* (1996) "Frequency of recurrent BRCA1 and BRCA2 mutations in Ashkenazi Jewish breast cancer families." *Nat Med.*; 2 (11): 1179-83.

Tupler R, Perini G, Green MR (2001) "Expressing the human genome." *Nature*; 409 (6822): 832-3.

Venkitaraman AR (2002) "Cancer susceptibility and the functions of BRCA1 and BRCA2." *Cell*; 108 (2): 171-82.

Vogelstein B & Kinzler KW (2004) "Cancer genes and the pathways they control" *Nat Med.* 10 (8): 789-99.

Wicking C, Simms LA, Evans T, Walsh M, Chawengsaksophak K, Beck F, Chenevix-Trench G, Young J, Jass J, Leggett B, Wainwright B (1998) "CDX2, a human homologue of *Drosophila* caudal, is mutated in both alleles in a replication error positive colorectal cancer." *Oncogene*; 17 (5): 657-9.

Wiseman BS & Werb Z (2002) "Stromal effects on mammary gland development and breast cancer." *Science*; 296 (5570): 1046-9.

## **ANEXOS**

## TABELA I

### Primers de amplificação do cDNA do SIX4

Exão 1	Amplificação 1	<b>1F: M13-GAGTGAGGGAGGGAGGAAAA</b> <b>1R: ctgcag tgcctcgcacac</b>
	Amplificação 2	<b>2F: M13- ACTccactcggaacttctgg</b> <b>2R: Cagccggtacttgctacgg</b>
	Amplificação 3	<b>3F: M13- gtacaaggecgctacacc</b> <b>3R: ggggaggaggaggaaagt</b>
Exão 2	Amplificação 4	<b>1F: M13- ATCCCCCTGACCTTCTGAAC</b> <b>1F: CACGTCCTGGGAAGTAGACC</b>
	Amplificação 5	<b>2F: M13- AGCAATGGTATATCCATGACTGA</b> <b>2R: GAACACATTTGCTGGTGCAT</b>
Exão 3	Amplificação 6	<b>1F: M13- tccctctcacttccatcagtg</b> <b>1R: gggactgagagaaaaattgtga</b>
	Amplificação 7	<b>2F: M13- atccactggcctcctcatta</b> <b>2R: tgcagaaccaaacgatgttc</b>
	Amplificação 8	<b>3F: M13- cccctcccagcagtacatc</b> <b>3R: atgccaaaaagaggtgagga</b>

**Primer universal: M13:GTAAAACGACGGCCAGT**

## TABELA II

### Primers de amplificação do cDNA do SIX4

Segmento 1	<b>07-79_</b> *1F: ccc gcc acc ATG TCC TCT TCC TCC CCC AC <b>07-80_</b> 1R: GCT CTC GAG GAT GCT GTA G
Segmento 2	<b>07-81_</b> 2F: CGA GCT CTA CAG CAT CCT C <b>07-82_</b> 2R: GGC TCC ATA TGA CTG GAA AG
Segmento 3	<b>07-83_</b> 3F: ATG GCA TCA CCA ACC TCA G <b>07-84_</b> 3R: CCT TGT GAA GCT GCC ACT G
Segmento 4	<b>07-85_</b> 4F: GGA TTC TCT CCA CTG CAG C <b>07-86_</b> 4R: CTG AGT CAT AGG GAC TGA C
Segmento 5	<b>07-87_</b> 5F: CTG CTCCTTATT ACTGGTC <b>07-88_</b> CCT TGG GGC TTC ATG AAA AG