

# **An Investigation of the Mechanism of PAX7 Mediated Oncogenesis via In Silico and In Vitro Biology**

**Maika Graceina Mitchell**

DISSERTATION.COM



Boca Raton

*An Investigation of the Mechanism of PAX7 Mediated Oncogenesis  
via In Silico and In Vitro Biology*

Copyright © 2007 Maika Graceina Mitchell

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the publisher.

Dissertation.com  
Boca Raton, Florida  
USA • 2008

ISBN-10: 1-59942-671-4  
ISBN-13: 978-1-59942-671-6

**An Investigation of the Mechanism of *PAX7* Mediated Oncogenesis via *In Silico* and *In Vitro* Biology**

Maika Graceina Mitchell

This proposal is presented in fulfillment of the requirements for the degree of

Doctor of Philosophy (Interdisciplinary Studies)

Faculty of Computing, Health and Science

Edith Cowan University

December 2006

**Committee:**

Melanie Ziman, Ph.D., Supervisor

Diane Tabarini, Ph.D. Co-Supervisor

## ABSTRACT

The *Pax* gene family appears to have evolved by a combination of gene duplication and / or genome duplication events over a long period of evolutionary time. The highly conserved paired box sequence within the *Pax* genes encodes a paired DNA binding domain, indicating that the Pax proteins are transcription factors which bind and regulate downstream target genes. Nine *Pax* genes (*Pax1* - *Pax9*) listed in the National Center for Biotechnology Information (NCBI) database, contain this motif.

Some members of the *Pax* family, which includes *Pax3* and *Pax7*, encode a second DNA binding domain of the paired-type homeodomain (HD) class. *Pax3* and *Pax7* are closely related paired box family members specifically expressed in the dorsal neural tube and the developing somites and in proliferating and migrating neural crest cells where they are implicated in early neural and myogenic development, and are required for development of specific myogenic, neurogenic and neural crest cell lineages. *Pax3* and *Pax7* genes are also found aberrantly expressed in tumors arising from these cell lineages.

The aim of the research was to analyze the molecular mechanism or mechanisms of *PAX7* mediated oncogenesis. This was achieved by:

Systematic searches for *cis* conserved sequences within *PAX7* intronic regions, which may be implicated in aberrant *PAX7* expression in tumours;

Comparison of conserved putative *cis* elements in human, mouse, chick and zebrafish *PAX7/Pax7* homologues to identify the most conserved regions as these are more likely to be functional *cis* acting regions;

Comparison of the conserved putative *cis* elements in human *PAX7* with those of other human *PAX* genes with a view to determining the most conserved regions, to assist with identification of likely functional *cis* acting regions;

Analysis of the oncogenic potential of *trans* factors likely to bind to identified *PAX7* putative *cis* elements;

Identification of polymorphisms within or close to identified putative *cis* elements so as to provide markers for genome-wide association mapping studies to identify Rhabdomyosarcoma susceptibility loci of *Homo sapiens*

The computational methodologies included but were not limited to systematic compilations of biological and computational results from various sources and evaluations of original experimental data with biocomputational tools and *in vitro* studies.

From our studies we identified a region in intron 8 of *PAX7* that is also found in intron 23 of the *NF-1* gene as well as in the alternative intron 10 of *PAX3*. This sequence appears to contain regulatory sequences that are conserved in all three genes and thus it seems probable that transcription factors and/or spliceosomes that bind to this region would act similarly on all three genes.

Regions of LOH, usually arising as a result of either hemizygous deletion or gene conversion events, are typically defined as stretches of chromosomal areas where all heterozygous and thereby informative alleles are rendered homozygous in the cancer. This classical definition assumes that all data points are accurately identified and that all polymorphic alleles are mapped correctly within the genome. In this project we used *in silico* biology to identify additional polymorphic sites that may provide information on LOH in future studies.

Interestingly there were no changes in SNP frequency observed in the ARMS samples relative to the expected allele frequency at the selected SNPs. Since only a few SNP sites were investigated in a very few samples for this thesis, additional SNP analysis at other identified sites may reveal significant changes in allele frequency and LOH in ARMS patients.

#### NOTE

The *Pax* gene encodes the murine Pax protein

The *PAX* gene encodes the human PAX protein

As most developmental biology studies have been performed on mice, the mouse form of Pax will be used unless specifically referring to human in which case PAX will be used.

## DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

Incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education.

Contain any material previously published or written by another person except where due reference is made in the text; or

Contain any defamatory material.

I also grant permission for the Library at Edith Cowan University to make duplicate copies of my thesis as required.

Signature: *MaiKa Graceina Mitchell*

Date: 12/11/2006

## **Dedication**

To my dear, loving husband, Carlton Mitchell, my son Cody, and my daughter Carmen for their undying support and encouragement in my continued pursuit of higher education. I must thank Dr. Melanie Ziman for being patient with me and tackling such a difficult pursuit as supervising me here in America from Australia. I also could not have completed the research portion of my doctoral degree without the permission of my Lab Manager Dr. Diane Tabarini-Ziff who also dedicated her precious time and provided much technical support and use of the lab facilities.

## **List of Abbreviations**

**NCBI**, National Center for Biotechnology Information  
**HD**, homeodomain  
**ARMS**, Alveolar Rhabdomyosarcoma  
**ERMS**, Embryonal Rhabdomyosarcoma  
**Krd**, kidney & retinal defects  
**BSAP**, B-cell lineage specific activator protein  
**TG**, thyroglobulin  
**TPO**, thyroperoxidase  
**HTH**, helix-turn-helix  
**Prd**, Paired  
**Gsb-p**, Gooseberry-proximal  
**MI**, Microsatellite Instability  
**RMS**, Rhabdomyosarcomas  
**HS**, Heat shock  
**RACE**, Rapid Amplification of c-terminus Ends  
**INSD**, The International Nucleotide Sequence Database Collaboration  
**EST**, expressed sequence tags  
**TrEMBL**, Translated EMBL Nucleotide Sequence Data Library  
**NLM**, National Library of Medicine  
**EMBL**, European Molecular Biology Laboratory  
**UniProt**, Universal Protein Resource  
**PDB**, Protein Data Bank  
**MSD**, The Macromolecular Structure Database  
**IVIAME**, Minimum Information about a Microarray Experiment  
**GEO**, Gene Expression Omnibus  
**GO**, Gene Ontology  
**DAG**, directed acyclic graph  
**SNPs**, Single-nucleotide polymorphisms  
**VNTRs**, Variable number of tandem repeats  
**HMM**, Hidden Markov Models  
**AFLP**, Amplified Fragment Length Polymorphisms  
**MLPA**, Multiplex Ligation-dependent Probe Amplification  
**T-RFLP**, Terminal-Restriction Fragment Length Polymorphism  
**EPCLUST**, Expression Profile data CLUSTERing and analysis  
**MICER**, Mutagenic Insertion and Chromosome Engineering Resource  
**ORFs**, Open reading frames



### List of Abbreviations continued...

**ml**, milliliter

**ul**, microliter

**nm**, nanometers

**ddH<sub>2</sub>O**, double distilled water

**PCR**, polymerase chain reaction

**AQ**, allele quantification

**PSQ**, Pyrosequencer

**NF-1**, neurofibromatosis factor 1

**TSS**, transcription start site

**bp**, base pair

**TF**, Transcription Factor

**SSLP**, Simple sequence length polymorphisms

**nt**, nucleotide

**LOH**, Loss of heterozygosity

**CNP**, Copy Number Polymorphism

**5'**, five prime end of DNA strand

**3'**, three prime end of DNA strand

**FKHR**, Fork Head Region

**MAS**, Marker Assisted Selection

### **Lists of figures and tables for chapter thesis ( excluding papers 1, 2, and 3)**

PAX1	Page 2
PAX9	Page 3
PAX2	Page 4
PAX5	Page 4
PAX8	Page 5
PAX3	Page 6
PAX7	Page 7
PAX4	Page 8
PAX6	Page 8
Figure 1. The paired domain of Pax-3 and Pax-7	Page 11
Figure 2. A density-modified MI (Microsatellite instability) map	Page 12
Figure 3. Pax paired domain-DNA complex	Page 13
Figure 4. Predicted structure of the Paired and homeodomain of Pax proteins	Page 15
Figure 5. A dorsal view of Hamburger and Hamilton stage 17 chick embryos	Page 17
Figure 6. Conserved locations of regulatory promoter elements	Page 23
Figure 7. Depiction of CAAT box, TATA box and GC box	Page 24
Figure 8. This HS-40 $\alpha$ -globin regulatory site	Page 25
Figure 9. Promoter region of human PAX7	Page 27

### **Lists of figures and tables for Paper 1**

Table 1. CLC Gene Workbench v.1.0.1. Pattern Discovery Search	Page 51
Figure 1. <i>Cis</i> regulatory region predicted by CLC Gene Workbench,	Page 51
Table 1. CLC Gene Workbench v.1.0.1. Pattern Discovery Search	Page 52
Figure 2. No <i>cis</i> element predicted, but 1 pattern found within 400-500 bps	Page 52
Table 2a-3b. CLC Gene Workbench v.1.0.1. Pattern	Page 52-53
Figure 3. <i>Cis</i> regulatory region predicted	Page 54
Table 5a-5f. CLC Gene Workbench v.1.0.1. Pattern Discovery	Page 55-56
Figure 5. <i>Cis</i> regulatory region predicted	Page 56
Table 6. CLC Gene Workbench v.1.0.1. Pattern Discovery Search	Page 57
Figure 6. <i>Cis</i> regulatory region predicted	Page 57
Table 7. CLC Gene Workbench v.1.0.1. Pattern Discovery Search	Page 58
Figure 7. <i>Cis</i> regulatory region predicted	Page 58
Table 8a-8e. CLC Gene Workbench v.1.0.1. Pattern Discovery Search	Page 58-59
Figure 8. <i>Cis</i> regulatory region predicted	Page 60
Table 9. Summary of most likely <i>cis</i> regulatory sequences	Page 60
Table 10. Summary of most likely <i>cis</i> regulatory sequences predicted	Page 60-61
Figure 9. Conserved sequence in intron 8 of PAX7	Page 61
Figure10. Conserved sequence in intron 8 of PAX7	Page 62

### **Lists of figures and tables for Paper 2**

Table 1. Primers	Page 68
Figure 2. Agarose Gel electrophoresis showing DNA isolated from volunteer buccal swabs, ERMS cell line and ARMS patient DNA	Page 70
Figure 3. Fragment analysis results for ARMS Patient #1, Primer 5 with electropherogram and peak table	Page 70
Figure 4. Fragment analysis results for ARMS Patient #2, Primer 5 with electropherogram and peak table	Page 71
Figure 5. Fragment analysis results for ARMS Patient 31, Primer 5 with electropherogram and peak table	Page 71
Figure 6. Fragment analysis results for ARMS Patient #4, Primer 5 with electropherogram and peak table	Page 71
Figure 7. Fragment analysis results for ARMS Patient #5, Primer 5 with electropherogram and peak table	Page 72
Figure 8 & Figure 9. Fragment analysis results for ERMS cell line(ATCC#: CCL 136 ) and control sample, with Primer 5 showing electropherograms and peak tables	Page 72

Figure 10. Fragment analysis results for ARMS Patient #1, Primer 16 with electropherogram and peak table	Page 73
Figure 11. Fragment analysis results for ARMS Patient #2, Primer 16 with electropherogram and peak table.	Page 73
Figure 12. Fragment analysis results for ARMS Patient #3, Primer 16 with electropherogram and peak table	Page 73
Figure 13. Fragment analysis results for ARMS Patient #4, Primer 16 with electropherogram and peak table	Page 74
Figure 14. Fragment analysis results for ARMS Patient #5, Primer 16 with electropherogram and peak table	Page 74
Figure 15. Fragment analysis results for ERMS Cell line (ATCC#: CCL 136), Primer 16 with electropherogram and peak table	Page 74
Figure 16. Fragment analysis results for CONTROL sample (genomic DNA) with Primer 16 with electropherogram and peak table.	Page 75
Figure 17. Fragment analysis results for ARMS Patient #1, Primer 37 with electropherogram and peak table.	Page 75
Figure 18. Fragment analysis results for ARMS Patient #2, Primer 37 with electropherogram and peak table.	Page 75
Figure 19. Fragment analysis results for ARMS Patient #3, Primer 37 with electropherogram and peak table	Page 76
Figure 20. Fragment analysis results for ARMS Patient #4, Primer 37 with electropherogram and peak table.	Page 76
Figure 21. Fragment analysis results for ARMS Patient #5, Primer 37 with electropherogram and peak table.	Page 76
Figure 22. Fragment analysis results for CONTROL sample (genomic DNA) with Primer 37 with electropherogram and peak table.	Page 77
Figure 23. Fragment analysis results for ERMS Cell line (ATCC#: CCL 136), Primer37 with electropherogram and peak table.	Page 77

### **Lists of figures and tables for Paper 3**

Figure 1. Cis element (GGGGATGGG indicated in teal; Mitchell and Ziman 2006) found in Human PAX7 intron eight located near SNP rs742074 (Y = IUPAC code for (C/T)).	Page 91
Figure 2. Cis element CTCCTCCC indicated in pink (Mitchell and Ziman 2006) found in Human PAX7 intron 8 located before SNP rs735630 (M = IUPAC code for C/A).	Page 92
Figure 3. Multiple alignment of PAX7 transcript variants.	Page 92
Figure 4. Invitrogen™ 4% precast Agarose Gel electrophoresis showing serial dilution for DNA isolated from volunteer buccal swab.	Page 93
Figure 5. The results of the Biotage© Pyrosequencing™ PSQ HS 96A & Sanger sequencing for Primers 23, 24 & 40 for ARMS, and Control Samples.	Page 93-95

<u>TABLE OF CONTENTS</u>	<u>PAGE #</u>
<b>CHAPTER ONE: PAX GENES/ PROTEINS</b> .....	<b>1</b>
<b>A.THE PAX FAMILY</b> .....	<b>1</b>
1. The PAX1/PAX9 Subfamily.....	<b>2</b>
2. The PAX2/PAX5/PAX8 Subfamily.....	<b>3</b>
3. The PAX3/PAX7 Subfamily.....	<b>5</b>
4. The PAX4/PAX6 Subfamily.....	<b>7</b>
<b>B. PAX 7 PROTEIN FUNCTION</b>	
1. DNA Recognition Mediated by the Paired Domain.....	<b>10</b>
2. DNA Recognition Mediated by the Homeodomain.....	<b>14</b>
<b>C. FUNCTION OF PAX7 DURING EMBRYOGENESIS</b>	
1. Neurulation.....	<b>16</b>
2. Pax7 in the neural tube, neural crest cells and the dermamyotome of the somites.....	<b>17</b>
3. Pax7 in developing and adult muscle tissues.....	<b>18</b>
<b>D. RHABDOMYOSARCOMA</b>	
1. Alveolar Rhabdomyosarcoma (ARMS).....	<b>19</b>
2. Embryonal Rhabdomyosarcoma (ERMS).....	<b>19</b>
<b>CHAPTER TWO: THE REGULATORY REGIONS RESPONSIBLE FOR PAX7 EXPRESSION</b> .....	<b>21</b>
<b>A. IDENTIFICATION OF BINDING SITES USING <i>IN SILICO</i> BIOLIOGY</b>	
<b><i>IN SILICO</i> BIOLIOGY METHODS</b> .....	<b>21</b>
<b>B. PUTATIVE CIS ELEMENTS OF GENE</b> .....	<b>23</b>
- TATA Box.....	<b>23</b>
- GC box.....	<b>23</b>

- CAAT box.....	23
- Other <i>cis</i> -elements - activators and repressors.....	25
<b>C. KNOWN PUTATIVE CIS ELEMENTS RESPONSIBLE FOR PAX7 EXPRESSION.....</b>	<b>25</b>
<b>CHAPTER THREE: PAX7 MEDIATED ONCOGENIC PATHWAY: ANALYSIS FACILITATED BY <i>IN SILICO</i> BIOLOGY.....</b>	<b>28</b>
<b>A. AIMS OF PROJECT -</b>	
➤ OVERALL AIM:	
To analyze the molecular mechanism or mechanisms of <i>PAX7</i> mediated oncogenesis.....	28
➤ Aim 1.	
To identify <i>cis</i> -regulatory sequences in intronic regions of <i>PAX7</i> ( <i>Homo sapiens</i> ).....	28
➤ Aim 2	
To compare putative <i>cis</i> elements in human, mouse, chick and zebrafish <i>PAX7/Pax7</i> homologues to identify conserved regulatory sequences.....	28
➤ Aim 3	
To compare putative <i>cis</i> elements in <i>PAX7</i> ( <i>Homo sapiens</i> ) with those in other <i>PAX</i> genes ( <i>PAX1-9</i> ).....	28
➤ Aim 4	
To analyse the oncogenic potential of <i>trans</i> factors likely to bind to those putative <i>cis</i> elements identified in the intronic regions of <i>PAX7</i> .....	29
➤ Aim 5	
To identify polymorphism(s) in intronic regulatory sequences of <i>PAX7</i> .....	29

## **B. THEORETICAL FRAMEWORK**

The use of *In Silico* Biology and *in vitro* experiments to detect *cis* and *trans* factors that aberrantly regulate *PAX7* expression in Rhabdomyosarcoma.....30

**C. MATERIALS AND METHODS.....39**

**D. JOURNAL ARTICLES 1-3.....47**

1. Investigation into Discovery of Putative *cis* elements within the Intronic Regions of Human *PAX7*.....47

2. Loss of Heterozygosity Analysis at selected Single Nucleotide Polymorphism Sites in the Intronic Regions of *PAX7* via *In Silico* Biology & Microsatellite Analysis.....66

3. Single Nucleotide Polymorphisms, Present in the Fusion Gene of the Intronic *Cis* Regulatory Regions of *PAX7*, Detected (observed) in Rhabdomyosarcoma via *In Silico* Biology and Pyrosequencing.....82

**CHAPTER FOUR: *IN SILICO* RESULTS.....99**

**CHAPTER FIVE: DISCUSSION & CONCLUSION.....127**

**CHAPTER SIX: FUTURE RESEARCH INITIATIVES BASED ON RESULTS FOUND IN THIS THESIS.....128**

**REFERENCES.....129**

**APPENDICES.....139**

## CHAPTER ONE: *PAX* GENES/ *PAX* PROTEINS

### A. THE *PAX* FAMILY

*Pax* genes derive their name from the Paired box gene region which encodes a highly conserved Paired DNA binding domain. Paired domains are found in all members of the **PAX** family. The *PAX* gene family appears to have evolved by a combination of gene duplication and / or genome duplication events over a long period of evolutionary time. In man and mouse nine *PAX/Pax* genes have been found. Each member of the *PAX* gene family is expressed in a spatially and temporally restricted pattern during embryogenesis. There are four classes of *PAX* genes based not only on sequence but on genomic organization. Genes within a given class have intron/exon boundaries and encoding regions in common. Several *PAX* genes also encode an octapeptide and a full or partial paired type homeodomain. A proline-rich acidic region at the COOH terminus is identified as the transactivation domain for *PAX* proteins.

<b><i>PAX1 &amp; 9:</i></b>	have a paired box with no introns and an octapeptide encoding region
<b><i>PAX2, 5 &amp; 8:</i></b>	have a paired box, a homeobox and an octapeptide encoding region
<b><i>PAX3 &amp; 7:</i></b>	have a paired box, a homeobox and an octapeptide encoding region
<b><i>PAX4 &amp; 6:</i></b>	have a paired box and a homeobox

Mouse *Pax* genes are expressed in a distinct pattern throughout embryogenesis. Generally *Pax* genes which have both a paired box and a homeobox, are expressed earlier (Erickson et al., 1993). Early in development, expression is observed in mitotically active cells, whereas at later developmental stages, *Pax* genes are expressed in lineage restricted cells. *Pax1* and *9* are expressed only in mesodermally derived tissue. The other *Pax* genes are expressed in the ectoderm.

*Pax* gene products are thought to function primarily by binding to enhancer DNA sequences and modifying the transcriptional activity of bound downstream target genes (Chi and Epstein, 2002).

Their importance as developmental genes is highlighted by the corresponding mutated phenotypes. In the heterozygote, the mutated *Pax* gene is semi-dominant, whereas homozygote mice do not generally survive to birth. The dosage of these genes therefore is critical and the exact phenotype produced in the mutant is a combination of genetic and environmental factors.

## 1. The PAX1/PAX9 Subfamily

The Pax1 and Pax9 proteins, categorized as the Group I subfamily, are proteins with the paired domain and octapeptide but without a homeodomain (Ogasawara et al., 1999). Pax1 and Pax9 are present in the sclerotome and are required for proper formation of the vertebral column (Rodrigo et al., 2004).

### PAX1/Pax1

**PAX1: paired box gene 1; Chromosomal Location: 20p11.2** (HUMAN); molecular type=genomic DNA



**Pax1: paired box gene 1; Chromosomal Location: 2** (MOUSE) ; development stage=8.5 day embryo



Expression of *Pax1* mRNA in the embryonic thymus has also been reported (Balling et al., 1996). Expression starts in the early endodermal epithelium lining the foregut region and includes the epithelium of the third pharyngeal pouch, a structure giving rise to part of the thymus epithelium. In early stages of thymus development a large proportion of thymus cells express *Pax1*. With increasing age, the proportion of *Pax1*-expressing cells is reduced and in the adult mouse only a small fraction of cortical thymic stromal cells retains strong *Pax1* expression. Expression of *Pax1* in thymus epithelium is necessary for establishing the thymus microenvironment required for normal T cell maturation (Wallin et al., 1996).

**PAX1 Human Mutations:** Klippel-Feil syndrome associated with sacral agenesis

**Pax1 Mouse Mutations:** spinal defect (*undulated*)



## PAX9/Pax9

**PAX9**: paired box gene 9; Chromosomal Location: 14q12-q13.1(HUMAN); tissue type=Lung, small cell carcinoma.



**Pax9**: paired box gene 9; Chromosomal Location: 12 (MOUSE);development stage=day 11.



*Pax9* is expressed in the pharyngeal pouch, vertebral column, tail (mouse), head and limbs. *Pax9* is paralogous to *Pax1*. PAX9 has been associated with dominantly inherited forms of human tooth agenesis that mainly involves posterior teeth (Frazier-Bowers et al., 2002).

**PAX9 Human Mutations:** selective tooth agenesis

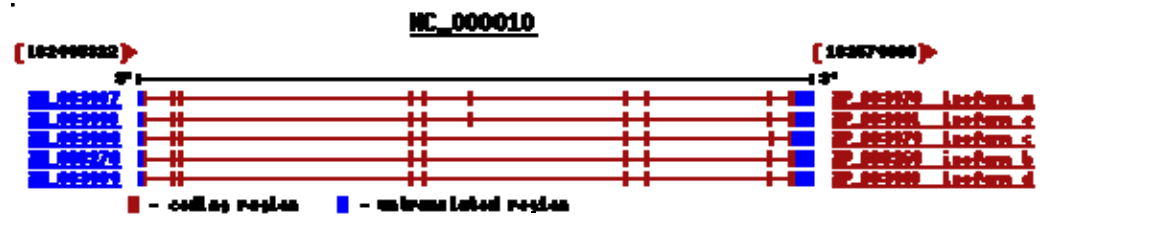
**Pax9 Mouse Mutations:** Tooth development is arrested at the bud stage; homozygote knockouts have secondary cleft palate and other abnormalities in craniofacial bones and cartilage.

## 2. The PAX2/PAX5/PAX8 Subfamily

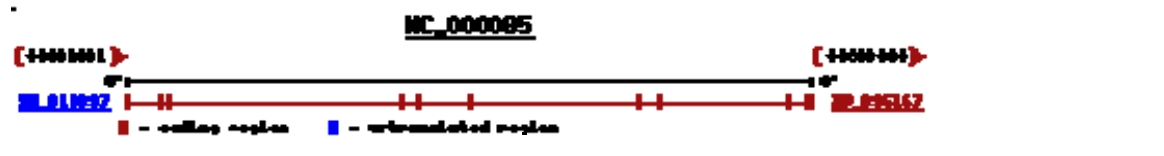
In vertebrates, the *PAX2*, *PAX5* and *PAX8* genes have been grouped into a common subfamily based on their sequence similarity and expression pattern. PAX2/PAX5/PAX8 protein members share in common the paired domain, an octapeptide motif, and a partial homeodomain. The *PAX2* gene is expressed during multiple stages of vertebrate nephrogenesis and when mutated, human genetic diseases of the genitourinary system arise (Majumdar et al., 2000).

## PAX2/Pax2

**PAX2**: Paired box gene 2; Chromosomal Location: 10q24.31 (HUMAN); tissue type=optic nerve; development stage=adult.



**Pax2**: Paired box gene 2; Chromosomal Location: 19 C3 (MOUSE); tissue type=brain; developmental stage=12 days



*PAX2* is expressed in the hindbrain and neural tube, optic stalk and vesicle, and during kidney organogenesis.

**PAX2 Human Mutations:** (optic nerve coloboma with renal disease)

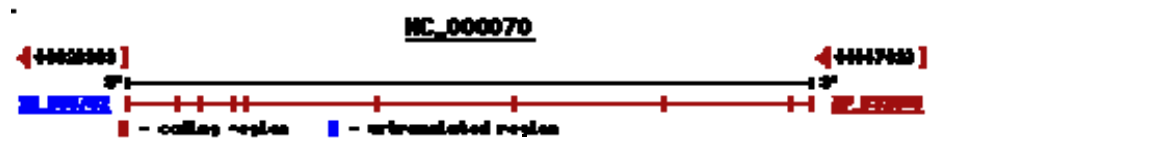
**Pax2 Mouse Mutations:** cause *Krd* (kidney & retinal defects)

## PAX5/Pax5

**PAX5**: Paired box gene 5 (B-cell lineage specific activator protein) Chromosomal Location: 9p13 (HUMAN); cell type=B-lymphocyte found in the developing CNS and adult testis.



**Pax5**: Paired box gene 5 Chromosomal Location: 4 B1 (MOUSE); cell type=B-lymphocyte found in the developing CNS and adult testis.



*Pax5* is expressed in fetal liver, B-lymphoid cells, mesencephalon and spinal cord. It is the B-cell lineage specific activator protein (BSAP) and controls expression of the CD19 gene (earliest B-lineage restricted cell surface antigen). *Pax5* (BSAP) acts both as a transcriptional activator and a repressor (Alexander et al., 2002). In addition to its role early in B cell differentiation, *Pax5* is also essential for later stages, when it influences the expression of many genes (Horcher et al., 2001).

**PAX5 Human Mutations:** possible link between *PAX5* and human primary immunodeficiencies (Vorechovsky et al. 1995)

**Pax5 Mouse Mutations:** cause *Krd*

### PAX8/Pax8

**PAX8:** paired box gene 8; Chromosomal Location: 2q12-q14 (HUMAN); tissue type=thyroid gland, fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid; development stage=fetal and adult



**Pax8:** paired box gene 8; Chromosomal Location: 2 B (MOUSE); tissue type=Kidney, normal, 5 month old male mouse.



Murine *Pax8* gene is expressed in the developing secretory system as well as in the developing and adult thyroid. This restricted expression pattern suggests involvement of the *Pax8* gene in morphogenesis of the above organs and prompted investigation of the *PAX8* gene in humans (Poleev, 1992). Human *PAX8* is present in both the thyroid and kidney and it transactivates two thyroid specific genes, *thyroglobulin* (*TG*) and *thyroperoxidase* (*TPO*).

**PAX8 Human Mutations:** Thyroid dysgenesis (congenital hypothyroidism); follicular carcinoma

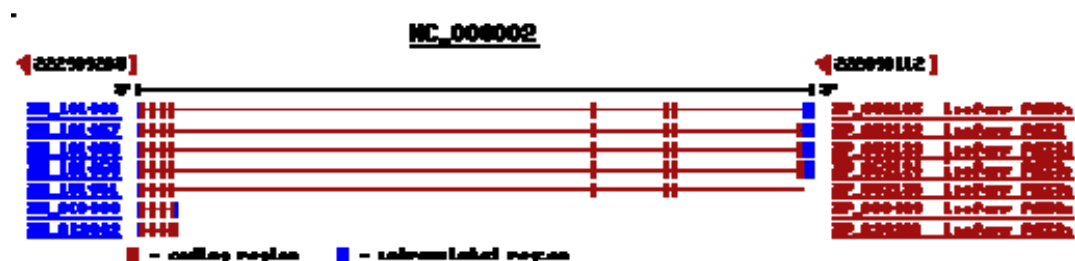
**Pax8 Mouse Mutations:** kidney malformations

### 3. The *PAX3/PAX7* Subfamily

*PAX3* and *PAX7* are closely related paired box family members expressed during early neural and myogenic development. *Pax3* and *Pax7* genes have been implicated in the development of specific myogenic and neurogenic cell lineages (Glaser et al., 1996; Relaix et al., 2004). Assay of *PAX3* and *PAX7* mRNA expression in embryonal rhabdomyosarcoma, neuroblastoma, Ewing's sarcoma, and melanoma cell lines revealed tumor-specific expression patterns that correspond to expression patterns in corresponding embryonic cell lineages (Barr et al., 1999; Zhang et al., 2003; Barr et al., 2005).

#### PAX3/Pax3

*PAX3*: Paired box gene 3; Chromosomal Location: 2q35-q37;2q35 (HUMAN);  
cell\_type=fibroblast, male.



*Pax3*: Paired box gene 3; Chromosomal Location: 1 C4 (MOUSE); tissue type=parthenogenote (the growth and development of an embryo or seed without fertilization by a male); development stage=9.5 days embryo.



*PAX3* is expressed in the neural tube, neural crest, dermomyotome & limb buds.

*PAX3* Human Mutation: *Waardenburg syndrome WS1* and *Klein- Waardenburg WS III* (pigmentary disturbances, dystopia canthorum, deafness in WS I plus limb abnormalities in WS III)

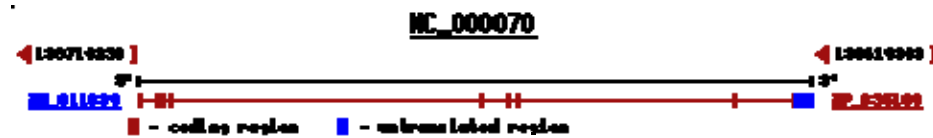
*Pax3* Mouse Mutation: *Spotch* mouse phenotype

## PAX7/Pax7

**PAX7:** Paired box gene 7; Chromosomal Location: 1p36.2-p36.12 (HUMAN); tissue type=alveolar rhabdomyosarcoma tumor, containing t(2;13); isolate=patient 282A.



**Pax7:** Paired box gene 7; Chromosomal Location: 4 E1 (MOUSE); tissue type=skeletal muscle.



*Pax7* is expressed in the developing nervous system; initially in the dorsal ventricular zone of the neural tube and later in the mesencephalon. *Pax7* is also expressed in the neural crest and the dermamyotome. *Pax7* is specifically expressed in satellite cells of skeletal muscle and is required for the specification of the satellite cell lineage (Seale et al., 2000).

**PAX7 Human Mutations:** alveolar rhabdomyosarcoma

**Pax7 Mouse Mutations:** Failure of caudal pharyngeal morphogenesis, small musculature, and limited muscle regeneration.

## 4. The PAX4/PAX6 Subfamily

The paired-homeodomain transcription factor Pax4 is present in the developing pancreas and along with Pax6 is required for normal development of endocrine cells. In the absence of Pax4, the numbers of insulin-producing  $\beta$  cells and somatostatin-producing  $\delta$  cells are drastically reduced, while the numbers of glucagon-producing  $\alpha$  cells are increased (Smith, 1999).

## PAX4/Pax4

**PAX4**: Paired box gene 4: Chromosomal Location: 7q22-qter (HUMAN); tissue type=placenta



**Pax4**: Paired box gene 4: Chromosomal Location: 6 A3.3 (MOUSE); cell type=pancreas.



**Pax4**, of all the nine members, has the most divergent paired domain of the family. Although the **Pax4** is essential for differentiation of insulin-producing beta-cells in the pancreas. **Pax4** has also been identified as a regulator of endocrine development, and has been shown to target gene promoters in an alpha-TC1.6 cell line (Frank et al., 2004).

## PAX6/Pax6

**PAX6**: Paired box gene 6: Chromosomal Location: 11p13 (HUMAN); tissue type=total brain; development stage=3 months old.



**Pax6**: Paired box gene 6: Chromosomal Location: 2 E3 (MOUSE); tissue type=embryo; development stage=day 8.5 post coitum, d 11.5 post coitum.



**Pax6** is expressed in the neural tube, in discrete areas of forebrain, hindbrain, eye and olfactory epithelium.

**PAX6** function was first identified through aniridia-associated null mutations. Since then, this transcription factor, which contains a paired domain and a homeodomain, has become a

paradigm, illustrating remarkable functional conservation in developmental pathways of the eye. The *Small eye* mutant mouse and *Drosophila Eyeless* have served as major model systems in defining the multistage roles for *Pax6* in eye and olfactory system development throughout evolution. The overt phenotypic consequences of heterozygous human and mouse *PAX6/Pax6* mutations were initially confined to the eye, with some interesting genotype–phenotype correlations being noted. Recently, structural and functional abnormalities in the olfactory system have been identified. Alterations in brain structure have also been documented, in line with the wider forebrain and cerebellar expression of *Pax6* (Azuma et al., 2005; van Heyningen, & Williamson, 2002). The broad *Pax6* expression pattern is controlled by a number of long-range control elements, and its importance is reflected in the severe homozygote phenotype. Upstream regulators and a multitude of downstream targets of *Pax6* have been identified, and its varied tissue-specific functions are emerging (Heyningen, et al., 2002).

**PAX6 Human Mutation:** *aniridia*

**Pax6 Mouse Mutation:** *smalleye*

**IN SUMMARY:** *Pax* genes are responsible for early cell lineage determination of many tissues and play a role in proliferation and maintenance of the undifferentiated cell state. Increasingly, their aberrant expression is associated with tumors arising from these specific cell types.

## CHAPTER ONE: PAX GENES/PROTEINS

### B. PAX 7 PROTEIN FUNCTIONS

#### 1. DNA Recognition Mediated by the Paired Domain

The paired domain is a 128 amino acid conserved domain which was originally found encoded by the *Drosophila* segmentation genes *paired* and *gooseberry*. The paired DNA-binding domain is strongly conserved from nematodes to mammals (Treisman et al., 1991; Mikkola et al., 1999). Proteins containing paired domains are transcription factors which bind to DNA (Xu et al. 1995; Fitzsimmons, et al., 2001; Wang, et al., 2005). The Paired domain, N-terminal motif is encoded by eight isoforms of Pax7. Eight other PAX transcription factors (PAX1, PAX2, PAX3, PAX4, PAX5, PAX6, PAX8) and associated isoforms, contain this motif.

Within the paired domain are two sub-domains (PAI and RED) which might act independently or may cooperate in the recognition of specific target gene promoter sequences *in vivo* (Vogan et al., 1996; Zhang et al., 2005). Each sub domain contains a DNA binding domain that consists of a set of three alpha-helices arranged in a helix, helix-turn-helix (HTH) motif. Within each of these HTH motifs, it is the third helix that contacts the specific DNA target sequence (Figures 1-4). Human *PAX7* cDNA, isolated from primary myoblasts and expressed *in vitro* (Schafer et al., 1994), was analyzed for its DNA-binding properties and was shown to bind DNA in a sequence-specific manner similar to that of the paralogous PAX3 protein (Schäfer et al., 1994; Du et al., 2004; Du et al., 2005).

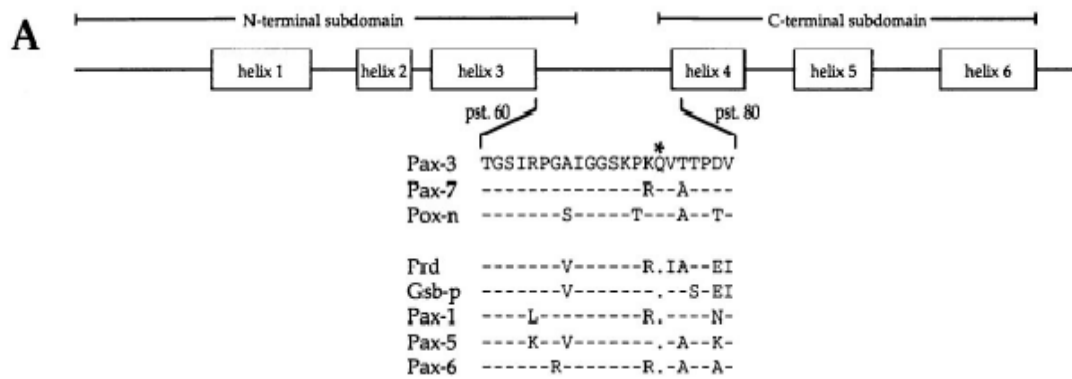


## Sequence of the paired domain in *Homo sapiens*

PAX7 PD Query: 1 GQGRVNQLGGVF INGRPLPNNIRL KIVEM AADGIRPCVISRQLRVSHGCVSKILNRYQET 60

PAX7 PD Query: 61 GSIRPGVIGGSKPRIATPEIENRIEEYKRSSPGMFSVEIREKLIREGVCDRSTAPSVSAI 120

PAX7 PD Query: 121 SRLVRGRD 128



**Figure 1.** The paired domain of Pax-3 and Pax-7. (A) Schematic representation of the paired domain. The N- and C-terminal sub domains are shown, with the six  $\alpha$ -helices indicated by rectangular boxes. Shown below are the sequences of the linker regions (residues 60 to 80) of Pax-3, Pax-7, and other *Drosophila* and mammalian paired-domain-containing proteins (24, 30). Identical residues are indicated by dashes. Abbreviations for the *Drosophila* proteins: Pox-n, Pox-neuro; Prd, Paired; Gsb-p, Gooseberry-proximal. The additional glutamine in alternate isoforms of Pax-3, Pax-7, and *Drosophila* Pox-n is indicated by an asterisk. pst., position. (Vogan et al., 1996).