A Tale of Two Transcription Factors:
The Interaction Between PAX8 and E2F1 in Cancer Cells

Jennifer E. Nyman

A thesis submitted for the degree of
Master of Science, Genetics
University of Otago
Dunedin, New Zealand
2012
Abstract

The *PAX* gene family encodes a group of transcription factors that play important roles during embryogenesis with mutations resulting in developmental abnormalities. Normally expression of these genes is downregulated prior to terminal differentiation with a select few tissues maintaining expression in the adult. Expression of *PAX* genes has also been observed in a variety of different cancers, although the functions of the aberrantly expressed *PAX* proteins have yet to be determined. Previous research has shown that knockdown of *PAX8* in cancer cell lines results in the downregulation of E2F1 mRNA and protein levels. *E2F1* is the major regulator of the cell cycle and its expression is often deregulated in cancer.

The overall aim of this study was to investigate the relationship between *PAX8* and *E2F1* in cancer cells. The first step was to determine the effects of *PAX8* knockdown and overexpression on the activity of the *E2F1* promoter, using two *E2F1* promoter constructs, pE2F1(-242) and pE2F1(-728). The results suggested that *PAX8* was involved in the regulation of the *E2F1* promoter as the activity levels of both *E2F1* constructs decreased following *PAX8* knockdown and increased following *PAX8* overexpression.

The next step was to determine if *PAX8* was regulating the *E2F1* promoter directly using chromatin immunoprecipitation (ChIP). Six potential binding sites were predicted in the region of the *E2F1* promoter contained in the longer of the two constructs, pE2F1(-728). Of these six predicted sites, *PAX8* was found to bind to site 3, thereby demonstrating that *PAX8* activates *E2F1* transcription by binding directly to the promoter.

Finally, deletion constructs and site-directed mutagenesis were employed in an attempt to validate the ChIP results as well as to determine if *PAX8* was able to bind to site 2, which was not included in the ChIP experiment as successful primers were not able to be designed. Unfortunately these experiments were
unsuccessful and thus no definitive conclusions could be made. Despite this, it was determined that PAX8 binds to and activates the $E2F1$ promoter in cancer cells. The involvement of PAX8 in the regulation of proliferation (by activating $E2F1$ transcription) reveals a potential function for PAX8 in cancer.
Acknowledgements

I would like to start off by thanking my supervisor, Prof. Mike Eccles, for giving me this amazing opportunity. The patience, guidance, time and effort that he gave me throughout this process, particularly during the write-up period, were greatly appreciated.

Next I would like to thank the other members of the Developmental Genetics and Pathology Group who were there throughout my stay: Lynn, Steve, Pam, Aaron, Cherie, Grace, Gagan, Bron and Chris. They welcomed me into the lab, showed me the ropes, and were willing to answer any and all questions. I would also like to thank the rest of the Pathology Department who were always ready to help in any way they could.

I am very grateful to all of the friends that I made while in New Zealand, especially Robert, Sarah, Stephan, and Carrie. They helped make Dunedin feel like home and could always be counted on to put a smile on my face, provide a pep talk, make me laugh or just listen.

Last, but not least, I would like to thank my family for their never-ending support, patience, and encouragement. I couldn’t have done it without them.
# Table of Contents

ABSTRACT II

ACKNOWLEDGEMENTS IV

TABLE OF CONTENTS V

LIST OF TABLES IX

LIST OF FIGURES X

LIST OF ABBREVIATIONS XII

CHAPTER 1 1

1.1 DEVELOPMENT, THE CELL CYCLE, AND CANCER 1
1.2 THE PAX GENE FAMILY 2
   1.2.1 PAX PROTEIN FUNCTIONAL DOMAINS 2
      1.2.1.1 Paired Box 2
      1.2.1.2 Homeodomain 4
      1.2.1.3 Octapeptide 4
   1.2.2 PAX GENE EXPRESSION 4
      1.2.2.1 Group I 5
      1.2.2.2 Group III 6
      1.2.2.3 Group IV 7
      1.2.2.4 Group II 8
   1.2.3 PAX REGULATION OF TRANSCRIPTION 12
      1.2.3.1 GROUP III 12
      1.2.3.2 GROUP IV 13
      1.2.3.3 GROUP II 14
   1.2.4 PAX AND CANCER 15
      1.2.4.1 CHROMOSOMAL TRANSLOCATIONS 15
      1.2.4.2 PERSISTENT EXPRESSION 16
1.3 THE E2F GENE FAMILY 17
   1.3.1 E2F GENE EXPRESSION DURING THE CELL CYCLE 18
   1.3.2 E2F GENES AND CANCER 19
   1.3.3 E2F1 EXPRESSION 20
1.4 BACKGROUND 22
1.5 AIM 22

CHAPTER 2 23

2.1 CELL LINES 23
2.2 CELL CULTURE 23
   2.2.1 MAINTENANCE 23
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.2</td>
<td>TRANSIENT TRANSFECTION</td>
<td>23</td>
</tr>
<tr>
<td>2.2.2.1</td>
<td>DNA Transfection</td>
<td>26</td>
</tr>
<tr>
<td>2.2.2.2</td>
<td>siRNA Transfection</td>
<td>26</td>
</tr>
<tr>
<td>2.2.2.3</td>
<td>siRNA and Plasmid DNA Co-transfection</td>
<td>27</td>
</tr>
<tr>
<td>2.2.3</td>
<td>REPORTER GENE ANALYSIS</td>
<td>27</td>
</tr>
<tr>
<td>2.2.3.1</td>
<td>Harvesting Cell Lysate</td>
<td>27</td>
</tr>
<tr>
<td>2.2.3.2</td>
<td>Luciferase Assay</td>
<td>28</td>
</tr>
<tr>
<td>2.2.3.3</td>
<td>β-galactosidase Assay</td>
<td>28</td>
</tr>
<tr>
<td>2.2.3.4</td>
<td>Data Analysis</td>
<td>28</td>
</tr>
<tr>
<td>2.2.3.1</td>
<td>DNA Transfection</td>
<td>26</td>
</tr>
<tr>
<td>2.2.3.2</td>
<td>siRNA Transfection</td>
<td>26</td>
</tr>
<tr>
<td>2.2.3.3</td>
<td>siRNA and Plasmid DNA Co-transfection</td>
<td>27</td>
</tr>
<tr>
<td>2.2.3.4</td>
<td>Data Analysis</td>
<td>28</td>
</tr>
<tr>
<td>2.3</td>
<td>PROTEIN EXPRESSION ANALYSIS</td>
<td>29</td>
</tr>
<tr>
<td>2.3.1</td>
<td>TOTAL PROTEIN ISOLATION</td>
<td>29</td>
</tr>
<tr>
<td>2.3.2</td>
<td>PROTEIN QUANTIFICATION</td>
<td>30</td>
</tr>
<tr>
<td>2.3.2.1</td>
<td>Bovine Serum Albumin Standard Curve</td>
<td>30</td>
</tr>
<tr>
<td>2.3.2.2</td>
<td>BCA Assay</td>
<td>31</td>
</tr>
<tr>
<td>2.3.3</td>
<td>WESTERN BLOT</td>
<td>31</td>
</tr>
<tr>
<td>2.3.3.1</td>
<td>Separating Gel</td>
<td>31</td>
</tr>
<tr>
<td>2.3.3.2</td>
<td>Stacking Gel</td>
<td>32</td>
</tr>
<tr>
<td>2.3.3.3</td>
<td>Sample Preparation</td>
<td>32</td>
</tr>
<tr>
<td>2.3.3.4</td>
<td>SDS-PAGE Gel</td>
<td>32</td>
</tr>
<tr>
<td>2.3.3.5</td>
<td>Transfer</td>
<td>32</td>
</tr>
<tr>
<td>2.3.3.6</td>
<td>Development</td>
<td>33</td>
</tr>
<tr>
<td>2.3.4</td>
<td>DENSITOMETRY</td>
<td>34</td>
</tr>
<tr>
<td>2.4</td>
<td>PRIMERS</td>
<td>35</td>
</tr>
<tr>
<td>2.5</td>
<td>CHROMATIN IMMUNOPRECIPITATION</td>
<td>36</td>
</tr>
<tr>
<td>2.5.1</td>
<td>BINDING SITE PREDICTION</td>
<td>36</td>
</tr>
<tr>
<td>2.5.2</td>
<td>CELLS</td>
<td>37</td>
</tr>
<tr>
<td>2.5.3</td>
<td>CROSSLINKING</td>
<td>37</td>
</tr>
<tr>
<td>2.5.4</td>
<td>SONICATION</td>
<td>37</td>
</tr>
<tr>
<td>2.5.5</td>
<td>PRE-CLEARING</td>
<td>38</td>
</tr>
<tr>
<td>2.5.6</td>
<td>IMMUNOPRECIPITATION</td>
<td>38</td>
</tr>
<tr>
<td>2.5.7</td>
<td>ChIP-DNA ELUTION</td>
<td>39</td>
</tr>
<tr>
<td>2.5.8</td>
<td>INPUT DNA ELUTION</td>
<td>39</td>
</tr>
<tr>
<td>2.5.9</td>
<td>QUANTITATIVE PCR</td>
<td>40</td>
</tr>
<tr>
<td>2.5.9.1</td>
<td>Amplification Efficiency</td>
<td>40</td>
</tr>
<tr>
<td>2.5.9.2</td>
<td>ChIP qPCR</td>
<td>40</td>
</tr>
<tr>
<td>2.5.9.3</td>
<td>Amplicon Sequencing</td>
<td>41</td>
</tr>
<tr>
<td>2.5.10</td>
<td>DATA ANALYSIS</td>
<td>41</td>
</tr>
<tr>
<td>2.6</td>
<td>DELETION CONSTRUCTS</td>
<td>42</td>
</tr>
<tr>
<td>2.6.1</td>
<td>PCR</td>
<td>42</td>
</tr>
<tr>
<td>2.6.2</td>
<td>AGAROSE GEL ELECTROPHORESIS</td>
<td>42</td>
</tr>
<tr>
<td>2.6.3</td>
<td>DNA EXTRACTION</td>
<td>42</td>
</tr>
<tr>
<td>2.6.4</td>
<td>PREPARATION OF INSERT AND VECTOR DNA</td>
<td>43</td>
</tr>
<tr>
<td>2.6.5</td>
<td>LIGATION</td>
<td>43</td>
</tr>
<tr>
<td>2.6.6</td>
<td>TRANSFORMATIONS</td>
<td>44</td>
</tr>
<tr>
<td>2.6.7</td>
<td>SINGLE COLONY PCR</td>
<td>44</td>
</tr>
<tr>
<td>2.6.8</td>
<td>MINIPREPARATION OF PLASMID DNA</td>
<td>44</td>
</tr>
<tr>
<td>2.6.9</td>
<td>SEQUENCING</td>
<td>45</td>
</tr>
<tr>
<td>2.6.10</td>
<td>MIDIPREPARATION OF PLASMID DNA</td>
<td>45</td>
</tr>
<tr>
<td>2.6.11</td>
<td>PROMOTER-REPORTER ASSAY</td>
<td>45</td>
</tr>
<tr>
<td>2.7</td>
<td>SITE-DIRECTED MUTAGENESIS</td>
<td>46</td>
</tr>
<tr>
<td>2.7.1</td>
<td>PRIMER DESIGN</td>
<td>46</td>
</tr>
<tr>
<td>2.7.2</td>
<td>PCR</td>
<td>46</td>
</tr>
<tr>
<td>2.7.3</td>
<td>CLONING</td>
<td>47</td>
</tr>
<tr>
<td>2.7.4</td>
<td>PROMOTER-REPORTER ASSAY</td>
<td>47</td>
</tr>
</tbody>
</table>
# CHAPTER 3

3.1 **INTRODUCTION**

3.2 **E2F1 PROMOTER CONSTRUCTS**

3.2.1 **SEQUENCE VALIDATION**

3.3 **PAX8 KNOCKDOWN**

3.3.1 **VALIDATION OF siRNA**

3.4 **PAX8 OVEREXPRESSION**

3.4.1 **VERIFICATION OF THE PAX8 EXPRESSION PLASMID**

3.5 **DISCUSSION**

# CHAPTER 4

4.1 **INTRODUCTION**

4.2 **BINDING SITE PREDICTION**

4.3 **QUANTITATIVE PCR**

4.3.1 **PRIMER DESIGN**

4.3.2 **REACTION CONDITIONS**

4.3.3 **AMPLIFICATION EFFICIENCY**

4.4 **CHROMATIN IMMUNOPRECIPITATION**

4.4.1 **CELLS**

4.4.2 **OPTIMIZATION OF CELL LYSIS AND SONICATION**

4.5 **CHIP-qPCR**

4.5.1 **AMPLICON SEQUENCING**

4.6 **DISCUSSION**

4.6.1 **CHIP**

4.6.2 **CHIP-qPCR AND SEQUENCING**

4.6.3 **PREDICTED BINDING SITES**

4.6.4 **SITE 2**

4.6.4.1 **qPCR**

4.6.4.2 **PAX8 Binding**

# CHAPTER 5

5.1 **INTRODUCTION**

5.2 **DELETION CONSTRUCTS**

5.2.1 **CLONING**

5.2.1.1 **Primers**

5.2.1.2 **Amplification**

5.2.1.3 **DNA Purification**

5.2.1.4 **Restriction Digests**

5.2.1.5 **Ligation and Transformation**

5.2.1.6 **Single Colony PCR**

5.2.1.7 **DNA Purification and Sequencing**

5.2.2 **PROMOTER-REPORTER ASSAY**

5.2.2.1 **PAX8 Activation**

5.2.2.2 **E2F1 Activation**

5.2.3 **DELETION CONSTRUCT ANALYSIS**

5.2.3.1 **Insert Orientation**

5.2.3.2 **Deletion Mutants**

5.2.3.3 **Plasmid DNA Concentration**

5.2.3.4 **DNA Preparation**

5.3 **SITE-DIRECTED MUTAGENESIS**
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>TABLE 1-1: GENETIC ALTERATIONS OF E2F1 IN HUMAN CANCERS</td>
<td>21</td>
</tr>
<tr>
<td>1-2</td>
<td>TABLE 2-1: CELL LINES USED IN THIS STUDY</td>
<td>24</td>
</tr>
<tr>
<td>1-3</td>
<td>TABLE 2-2: PLASMID DNA USED IN THIS STUDY</td>
<td>25</td>
</tr>
<tr>
<td>1-4</td>
<td>TABLE 2-3: SIRNAS USED IN THIS STUDY</td>
<td>26</td>
</tr>
<tr>
<td>1-5</td>
<td>TABLE 2-4: SAMPLE CALCULATIONS OF RELATIVE LUCIFERASE UNITS (RLU)</td>
<td>29</td>
</tr>
<tr>
<td>1-6</td>
<td>TABLE 2-5: BSA STANDARD PREPARATION</td>
<td>31</td>
</tr>
<tr>
<td>1-7</td>
<td>TABLE 2-6: ANTIBODIES USED FOR WESTERN BLOTTING</td>
<td>34</td>
</tr>
<tr>
<td>1-8</td>
<td>TABLE 2-7: EXPERIMENT-SPECIFIC PRIMERS USED IN THIS STUDY</td>
<td>35</td>
</tr>
<tr>
<td>2-1</td>
<td>TABLE 2-1: QUANTITIES OF DNA TRANSFECTED IN EACH SAMPLE FOLLOWING</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>SIRNA TREATMENT</td>
<td></td>
</tr>
<tr>
<td>2-2</td>
<td>TABLE 2-2: QUANTITIES OF DNA TRANSFECTED INTO HEK-293 CELLS TO</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>EXAMINE THE EFFECTS OF PAX8 OVEREXPRESSION ON THE E2F1 PROMOTER</td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>TABLE 4-1: CHIP TARGET SEQUENCES AND AMPLICON LENGTHS</td>
<td>64</td>
</tr>
<tr>
<td>2-4</td>
<td>TABLE 4-2: AMPLIFICATION EFFICIENCY OF EACH CHIP PRIMER SET</td>
<td>68</td>
</tr>
<tr>
<td>2-5</td>
<td>TABLE 4-3: CT VALUES OBTAINED FOLLOWING EACH OF FOUR BIOLOGICAL</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>REPLICATE QPCR EXPERIMENTS AMPLIFYING INPUT, PAX8 AND NO-ANTIBODY</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(NOAB) CONTROL DNA OBTAINED FROM CHIP USING PRIMERS FOR SITES 1, 3-6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AND CONTROL</td>
<td></td>
</tr>
<tr>
<td>2-6</td>
<td>TABLE 4-4: ASSAY SITE IP FOLD ENRICHMENT VALUES FOR EACH PREDICTED</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>BINDING SITE</td>
<td></td>
</tr>
<tr>
<td>2-7</td>
<td>TABLE 5-1: NUMBER OF COLONIES OBTAINED FROM TRANSFORMATIONS DONE</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>FOLLOWING LIGATION OF THE F5 AND F6 INSERTS TO THE PGL2-BASIC VECTOR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>USING TWO DIFFERENT LIGATION CONDITIONS</td>
<td></td>
</tr>
</tbody>
</table>
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGURE 1-1:</td>
<td>THE HUMAN PAX GENE FAMILY</td>
</tr>
<tr>
<td>FIGURE 1-2:</td>
<td>THE E2F TRANSCRIPTION FACTOR FAMILY</td>
</tr>
<tr>
<td>FIGURE 1-3:</td>
<td>E2F EXPRESSION DURING THE CELL CYCLE</td>
</tr>
<tr>
<td>FIGURE 2-1:</td>
<td>SITE-DIRECTED MUTAGENESIS REACTIONS</td>
</tr>
<tr>
<td>FIGURE 3-1:</td>
<td>E2F1 PROMOTER CONSTRUCTS</td>
</tr>
<tr>
<td>FIGURE 3-2:</td>
<td>E2F1 PROMOTER CONSTRUCT SEQUENCE ALIGNMENT</td>
</tr>
<tr>
<td>FIGURE 3-3:</td>
<td>E2F1 PROMOTER ACTIVITY FOLLOWING PAX8 KNOCKDOWN</td>
</tr>
<tr>
<td>FIGURE 3-4:</td>
<td>PAX8 EXPRESSION</td>
</tr>
<tr>
<td>FIGURE 3-5:</td>
<td>E2F1 PROMOTER ACTIVITY FOLLOWING PAX8 OVEREXPRESSION</td>
</tr>
<tr>
<td>FIGURE 3-6:</td>
<td>PAX8 EXPRESSION LEVELS</td>
</tr>
<tr>
<td>FIGURE 3-7:</td>
<td>DENSITOMETER ANALYSIS OF PAX8 EXPRESSION</td>
</tr>
<tr>
<td>FIGURE 3-8:</td>
<td>PAX8 DENSITY RATIOS</td>
</tr>
<tr>
<td>FIGURE 4-1:</td>
<td>PREDICTED PAX8 BINDING SITES IN THE E2F1 PROMOTER</td>
</tr>
<tr>
<td>FIGURE 4-2:</td>
<td>SCMATIC REPRESENTATION OF THE E2F1 PROMOTER CONSTRUCTS USED IN THIS STUDY IN RELATION TO THE E2F1 PROMOTER</td>
</tr>
<tr>
<td>FIGURE 4-3:</td>
<td>SCMATIC REPRESENTATION OF THE LOCATIONS OF THE REGIONS TO BE AMPLIFIED BY THE CHIP PRIMERS</td>
</tr>
<tr>
<td>FIGURE 4-4:</td>
<td>PCR PRODUCTS FOR SITES 3-6 AND THE CONTROL</td>
</tr>
<tr>
<td>FIGURE 4-5:</td>
<td>PCR PRODUCTS USING THE RE-DESIGNED PRIMERS FOR SITES 1 AND 2</td>
</tr>
<tr>
<td>FIGURE 4-6:</td>
<td>PCR PRODUCTS USING THE SECOND SET OF RE-DESIGNED PRIMERS FOR SITE 2</td>
</tr>
<tr>
<td>FIGURE 4-7:</td>
<td>CHIP PRIMER STANDARD CURVES</td>
</tr>
<tr>
<td>FIGURE 4-8:</td>
<td>SONICATION TRIAL – NP-40 LYSIS BUFFER</td>
</tr>
<tr>
<td>FIGURE 4-9:</td>
<td>SONICATION TRIAL – NP-40 LYSIS BUFFER VERSUS SDS LYSIS BUFFER</td>
</tr>
<tr>
<td>FIGURE 4-10:</td>
<td>SONICATION TRIAL – SDS LYSIS BUFFER VERSUS RIPA BUFFER</td>
</tr>
<tr>
<td>FIGURE 4-11:</td>
<td>SONICATION TRIAL – NUMBER AND LENGTH OF PULSES</td>
</tr>
<tr>
<td>FIGURE 4-12:</td>
<td>SONICATION TRIAL – NUMBER OF 20-SECOND PULSES</td>
</tr>
<tr>
<td>FIGURE 4-13:</td>
<td>CHIP SONICATION</td>
</tr>
<tr>
<td>FIGURE 4-14:</td>
<td>DISSOCIATION CURVES FOR THE SITE 2 AND SITE 3 QPCR PRODUCTS</td>
</tr>
<tr>
<td>FIGURE 4-15:</td>
<td>CHIP-QPCR SITE 2 PRODUCTS</td>
</tr>
<tr>
<td>FIGURE 4-16:</td>
<td>FOLD CHANGE OF EACH PREDICTED BINDING SITE RELATIVE TO THE CONTROL SITE</td>
</tr>
<tr>
<td>FIGURE 4-17:</td>
<td>FOLD ENRICHMENT OF EACH PREDICTED PAX8 BINDING SITE RELATIVE TO SITE 1</td>
</tr>
<tr>
<td>FIGURE 4-18:</td>
<td>CHIP-QPCR PRODUCTS TO BE SEQUENCED</td>
</tr>
<tr>
<td>FIGURE 4-19:</td>
<td>SEQUENCE ALIGNMENT OF CHIP-QPCR PRODUCTS WITH THE E2F1 REFERENCE SEQUENCE</td>
</tr>
<tr>
<td>FIGURE 4-20:</td>
<td>ALIGNMENT OF THE TRANSFAC CONSENSUS WITH THE PREDICTED PAX8 BINDING SITES</td>
</tr>
<tr>
<td>FIGURE 4-21:</td>
<td>PAX8 CONSENSUS SEQUENCE ALIGNMENT</td>
</tr>
<tr>
<td>FIGURE 4-22:</td>
<td>ALIGNMENT OF THE OPTIMAL PAX8 BINDING SITE CONSENSUS SEQUENCE WITH THE PREDICTED PAX8 BINDING SITES</td>
</tr>
<tr>
<td>FIGURE 4-23:</td>
<td>ALIGNMENT OF THE PREDICTED BINDING SITES WITH THE PUBLISHED PAX8 BINDING SITES</td>
</tr>
<tr>
<td>FIGURE 4-24:</td>
<td>THE OPTIMAL BINDING SITE SEQUENCES FOR PAX8 AND THE DROSOPHILA PAIRED (PRD) PROTEIN</td>
</tr>
<tr>
<td>FIGURE 5-1:</td>
<td>DELETION CONSTRUCTS</td>
</tr>
<tr>
<td>FIGURE 5-2:</td>
<td>DELETION CONSTRUCT</td>
</tr>
</tbody>
</table>
FIGURE 5-3: INSERT DNA FOR DELETION CONSTRUCTS
FIGURE 5-4: NHEI/HINDIII-DIGESTED INSERT DNA AND PGL2-BASIC VECTOR
FIGURE 5-5: SINGLE COLONY PCR PRODUCTS USED TO VERIFY PRESENCE OF INSERT DNA
FIGURE 5-6: PGL2-BASIC VECTOR DEPICTING THE LOCATIONS OF THE GL3F2 AND GL3R1 PRIMERS
FIGURE 5-7: E2F1 PROMOTER CONSTRUCT ACTIVITY FOLLOWING OVEREXPRESSION OF PAX8
FIGURE 5-8: THE ACTIVITY LEVELS OF THE PGL2-BASIC NEGATIVE CONTROL AND THE DELETION CONSTRUCTS, WITH (A) AND WITHOUT (B) THE DATA FROM THE LONGEST CONSTRUCT, F1.
FIGURE 5-10: THE ACTIVITY LEVELS OF THE DELETION CONSTRUCTS FOLLOWING CO-TRANSFECTION WITH E2F1, WITH (A) AND WITHOUT (B) THE F1 DATA.
FIGURE 5-11: DELETION CONSTRUCT INSERT ORIENTATION
FIGURE 5-12: ANALYSIS OF THE DELETION CONSTRUCTS BY RESTRICTION DIGEST.
FIGURE 5-13: PLASMID DNA ANALYSIS
FIGURE 5-14: DELETION CONSTRUCT PLASMID DNA ANALYSIS
FIGURE 5-15: DELETION CONSTRUCT ACTIVITY FOLLOWING OVEREXPRESSION OF PAX8
FIGURE 5-16: F1 DELETION CONSTRUCT ACTIVITY FOLLOWING CO-TRANSFECTION WITH VARYING COMBINATIONS OF PAX8 AND PCR3.1
FIGURE 5-17: DELETION CONSTRUCT PLASMID DNA PURIFIED BY MINI PREPARATION.
FIGURE 5-18: ACTIVITY OF MINIPREP PURIFIED PLASMID DNA FOLLOWING OVEREXPRESSION OF PAX8
FIGURE 5-19: PAX8 CONSENSUS SEQUENCE ALIGNED WITH THE PREDICTED SITE 3 SEQUENCE.
FIGURE 5-20: SITE DIRECTED MUTAGENESIS PRIMERS
FIGURE 5-21: PCR PRODUCTS OBTAINED USING MUTAGENIC PRIMER PAIRS.
FIGURE 5-22: DIGESTED INSERT AND VECTOR DNA.
FIGURE 5-23: PRODUCTS FROM SINGLE COLONY PCR OF MUTATED PLASMID DNA.
FIGURE 5-24: SINGLE COLONY PCR PRODUCTS.
FIGURE 5-25: MUTATED SITE 3 SEQUENCE ALIGNED WITH E2F1 PROMOTER REFERENCE SEQUENCE.
FIGURE 5-26: ACTIVITY LEVELS OF THREE DIFFERENT E2F1 PROMOTER CONSTRUCTS FOLLOWING OVEREXPRESSION OF PAX8 AND E2F1.
FIGURE 5-27: ANALYSIS OF PLASMID DNA QUALITY.
FIGURE 5-28: ANALYSIS OF PCMV5-PAX8 EXPRESSION VECTOR.
FIGURE 6-1: PAX8 BINDING SITE IN THE TG AND TPO PROMOTERS.
FIGURE 6-2: PAX8 BINDING SITES IN THE N/S PROMOTER.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-gal</td>
<td>beta-galactosidase</td>
</tr>
<tr>
<td>1°</td>
<td>primary</td>
</tr>
<tr>
<td>2°</td>
<td>secondary</td>
</tr>
<tr>
<td>Amp</td>
<td>ampicillin</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>APS</td>
<td>ammonium persulfate</td>
</tr>
<tr>
<td>ATCC</td>
<td>American Type Culture Collection</td>
</tr>
<tr>
<td>BCA</td>
<td>bicinchoninic acid</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>ChIP</td>
<td>chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CMV</td>
<td>cytomegalovirus</td>
</tr>
<tr>
<td>CO₂</td>
<td>carbon dioxide</td>
</tr>
<tr>
<td>cpk</td>
<td>congenital polycystic kidney</td>
</tr>
<tr>
<td>Ct</td>
<td>threshold cycle</td>
</tr>
<tr>
<td>CycA</td>
<td>cyclin A</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding domain</td>
</tr>
<tr>
<td>DF</td>
<td>dilution factor</td>
</tr>
<tr>
<td>dH₂O</td>
<td>distilled water</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco's Modified Eagle Medium</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTPs</td>
<td>deoxyribonucleotide triphosphates</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EMBL-EBI</td>
<td>European Molecular Biology Laboratory – European Bioinformatics Institute</td>
</tr>
<tr>
<td>EtBr</td>
<td>ethidium bromide</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>G0 phase</td>
<td>quiescent phase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>G1 phase</td>
<td>post-mitotic gap 1 phase</td>
</tr>
<tr>
<td>G2 phase</td>
<td>pre-mitotic gap 2 phase</td>
</tr>
<tr>
<td>GI</td>
<td>gastrointestinal</td>
</tr>
<tr>
<td>HCl</td>
<td>hydrochloric acid</td>
</tr>
<tr>
<td>HEK</td>
<td>human embryonic kidney</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
</tr>
<tr>
<td>HPD</td>
<td>homeodomain and paired domain binding site</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>IDT</td>
<td>Integrated DNA Technologies</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>IP</td>
<td>immunoprecipitation</td>
</tr>
<tr>
<td>Kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>LiCl</td>
<td>lithium chloride</td>
</tr>
<tr>
<td>Luc</td>
<td>luciferase</td>
</tr>
<tr>
<td>LZ</td>
<td>leucine zipper</td>
</tr>
<tr>
<td>M phase</td>
<td>mitotic phase</td>
</tr>
<tr>
<td>mA</td>
<td>milliampere</td>
</tr>
<tr>
<td>MB</td>
<td>marked box</td>
</tr>
<tr>
<td>MEM</td>
<td>minimum essential media</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>magnesium chloride</td>
</tr>
<tr>
<td>MQ-H₂O</td>
<td>Milli-Q water</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>NaOH</td>
<td>sodium hydroxide</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localization signal</td>
</tr>
<tr>
<td>noAB</td>
<td>no antibody</td>
</tr>
<tr>
<td>NP-40</td>
<td>Nonidet P-40</td>
</tr>
<tr>
<td>NTC</td>
<td>non-template control</td>
</tr>
<tr>
<td>oligo</td>
<td>oligonucleotide</td>
</tr>
<tr>
<td>ONPG</td>
<td>ortho-nitrophenyl-β-galactoside</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PBST</td>
<td>phosphate buffered saline + 0.1% Tween-20</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PD</td>
<td>paired domain</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonylfluoride</td>
</tr>
<tr>
<td>PWM</td>
<td>position weight matrix</td>
</tr>
<tr>
<td>qPCR</td>
<td>quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>RCC</td>
<td>renal cell carcinoma</td>
</tr>
<tr>
<td>Ref. seq.</td>
<td>reference sequence</td>
</tr>
<tr>
<td>RIPA</td>
<td>radioimmunoprecipitation</td>
</tr>
<tr>
<td>RLU</td>
<td>relative luciferase units</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>ROX</td>
<td>6-carboxy-X-rhodamine</td>
</tr>
<tr>
<td>RPCs</td>
<td>retinal progenitor cells</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>quantitative real-time PCR</td>
</tr>
<tr>
<td>S phase</td>
<td>synthesis phase</td>
</tr>
<tr>
<td>SCC</td>
<td>squamous cell carcinoma</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SEM</td>
<td>standard error of mean</td>
</tr>
<tr>
<td>siControl</td>
<td>ON-TARGETplus Non-targeting Pool siRNA</td>
</tr>
<tr>
<td>siPAX8</td>
<td>ON-TARGETplus SMARTpool PAX8 siRNA</td>
</tr>
<tr>
<td>siRNA</td>
<td>small interfering ribonucleic acid</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate-EDTA</td>
</tr>
<tr>
<td>TE</td>
<td>Tris EDTA</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N’-tetramethylethylenediamine</td>
</tr>
<tr>
<td>TFCs</td>
<td>thyroid follicular cells</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>UCSC</td>
<td>University of California Santa Cruz</td>
</tr>
<tr>
<td>UDG</td>
<td>uracil-DNA glycosylase</td>
</tr>
<tr>
<td>V</td>
<td>volt</td>
</tr>
</tbody>
</table>
1.1 Development, the Cell Cycle, and Cancer

During embryonic development there is a delicate balance between cell division and apoptosis that is essential for proper cell proliferation and differentiation. This balance is maintained in part by the activity of a combination of signal transduction and stress-response pathways (Nebert, 2002). Downstream of each pathway are transcription factors, which act to up- or down-regulate the expression of specific genes.

Oncogenesis is akin to embryogenesis but lacks the level of control normally observed due to a disruption of the balance between these processes. This can result from mutations occurring in any of the protein-encoding genes along these pathways, including the transcription factors, with defects accumulating in an increasing number of genes over time (Bartek and Lukas, 2001). Given that the activity of transcription factors is strictly controlled, events resulting in their aberrant expression or activation/inactivation as well as mutations and translocations can all contribute to tumourigenesis (Libermann and Zerbini, 2006). These modifications result in the deregulated expression of numerous genes, some of which are associated with tumour development and progression or cell cycle and growth-regulatory functions (Nebert, 2002). Hanahan and Weinberg (2000) identified six hallmark features of cancer, which include: sustained cell division, insensitivity to anti-growth signals, the means to avoid apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. These characteristics indicate that cancerous cells are able to escape from the normal constraints imposed upon them, allowing them to proliferate uninhibited.
Studies involving $HOX$ genes have led to the hypothesis that “oncology recapitulates ontogeny”, which suggests that genes expressed during embryogenesis but downregulated in the adult are re-expressed during carcinogenesis (Grier et al., 2005; Lappin et al., 2006). Consistent with this hypothesis is the $PAX$ gene family, which encodes a group of transcription factors that play an important role in organogenesis and, for the most part, are subsequently downregulated. These transcription factors have also been observed in a variety of cancers.

1.2 The $PAX$ Gene Family

The $PAX$ gene family encodes a group of transcription factors that play critical roles during the process of embryonic development. They are involved in the regulation of organogenesis and act by promoting cell proliferation and self-renewal, migration, differentiation, and resistance to apoptosis. Mutations in the $PAX$ genes or aberrant expression can result in numerous diseases and have also been observed in a variety of cancers (Chi and Epstein, 2002; Robson et al., 2006; Lang et al., 2007).

1.2.1 $PAX$ Protein Functional Domains

The $PAX$ proteins are characterized by the presence of the $Paired box$ domain. Nine members of the transcription factor family have been identified thus far in mammals (Dahl et al., 1997). These nine genes are further divided into four subgroups based on the presence or absence of two other functional domains: the octapeptide domain and the homeodomain (Figure 1.1) (Wang et al., 2008). An additional domain, the transactivation domain, is located at the carboxy-terminus of all of the $PAX$ proteins (Chi and Epstein, 2002).

1.2.1.1 Paired Box

The paired box, or paired, domain is 128 amino acids long, highly conserved, and involved in DNA-binding (Bopp et al., 1986; Treisman et al., 1991). The
domain actually consists of two separate subdomains, PAI and RED (PAI + RED = PAIRED) (Xu et al., 1995), joined by a flexible linker. Residues in both the amino-terminal PAI and carboxy-terminal RED subdomains are involved in DNA sequence recognition and several studies have found that the paired domains of PAX proteins from different subgroups are able to recognize similar DNA sequences (Czerny et al., 1993; Epstein et al., 1994a; Xu et al., 1995; Jun and Desplan, 1996). The occurrence of alternative splicing within the paired domain has been shown to alter the DNA recognition specificity of PAX3, PAX6, and PAX8 (Epstein et al., 1994b; Vogan et al., 1996; Kozmik et al., 1997).

![Figure 1-1: The human PAX gene family](image)

The PAX gene family is distinguished by the presence of the paired domain at the N'-terminal of each protein. The family members are further divided into groups based on the presence of additional structural domains, including the octapeptide domain and the homeodomain.
1.2.1.2 Homeodomain

In addition to the DNA-binding ability of the paired domain, several of the PAX proteins are also able to bind DNA via either a partial (Group II) or complete (Group III and IV) homeodomain (Figure 1.1) (Chi and Epstein, 2002). A helix-turn-helix motif 60 amino acids long, the homeodomain is less conserved among the PAX proteins than the paired domain. Most homeodomains recognize the TAAT core motif. The PAX homeodomains dimerize to form either hetero- or homodimers and bind to the palindromic DNA sequence 5’-TAAT(N)_{2,3}ATTA-3’ (Wilson et al., 1993). It is the combination of the paired domain and the homeodomain that determines the DNA-binding specificity of each individual PAX protein.

1.2.1.3 Octapeptide

Located between the paired domain and the homeodomain of the PAX proteins in Groups I-III is the octapeptide domain (Figure 1.1). The eight amino acid long motif is highly conserved and has the sequence (His/Tyr)Ser(Ile/Val)(Asn/Ser)Gly(Ile/Leu)LeuGly (Noll, 1993). There is evidence that the octapeptide domain is involved in transcriptional repression in PAX2, PAX5, and PAX8 (Lechner and Dressler, 1996; Eberhard et al., 2000). In the case of PAX5, members of the Groucho family of co-repressors inhibit its transcriptional activity by interacting with the octapeptide domain (Eberhard et al., 2000).

1.2.2 PAX Gene Expression

The PAX genes are expressed throughout organogenesis in a tissue-specific manner with genes of the same subgroup being expressed in similar patterns. While the expression of these genes is typically downregulated in adult tissues, there are some cell types that maintain expression.
1.2.2.1 Group I

*Pax1* and *Pax9* are highly homologous and both are expressed in the developing vertebral column and the third and fourth pharyngeal pouches and their derivatives, which include the thymus, parathyroid glands, ultimobranchial arches, Eustachian tubes and tonsils (Deutsch et al., 1988; Neubuser et al., 1995; Wallin et al., 1996; Hetzer-Egger et al., 2002; Rodrigo et al., 2003). More specifically, *Pax1* is involved in the development of the sternum, limb buds, scapula, pectoral and pelvic girdle (Timmons et al., 1994; Wallin et al., 1994; Neubuser et al., 1995), whereas *Pax9* plays a role in the development of the tail, head, limb buds, esophagus, teeth, and larynx of the mouse (Neubuser et al., 1995; Peters et al., 1998). *Pax1* and *Pax9* are co-expressed during the development of the thymus, although they perform different functions. *Pax9* is necessary earlier in the development process, whereas *Pax1* is required later on in development as it is involved in T-cell maturation (Lang et al., 2007). Expression of both *Pax1* and *Pax9* is maintained in the adult thymus, albeit at much lower levels than are seen during development (Lang et al., 2007). For example, *Pax1* is expressed in the thymic epithelial cells during development, whereas it is only expressed in a few cortical epithelial cells in the adult tissue (Wallin et al., 1996).

Three different mutations in the *Pax1* gene result in undulated mice, which lack Pax1 function and are characterized by defects in the axial skeleton (Balling et al., 1988; Wallin et al., 1994; Dietrich and Gruss, 1995). Particularly affected are the intervertebral discs, which are often missing entirely, and the intervertebral bodies, which may be malformed and smaller in size (Dahl et al., 1997). The thymus is also affected, showing a decrease in size as well as impaired T-cell maturation (Wallin et al., 1996). *PAX1* mutations have also been observed in human congenital disorders including Klippel-Feil syndrome and Jarcho-Levin syndrome, which are characterized by vertebral malformations (Bannykh et al., 2003; McGaughran et al., 2003). There is also evidence that Pax1/PAX1 may contribute to the pathogenesis of spina bifida in both mice and humans (Helwig et al., 1995; Hol et al., 1996). In the case of *PAX9*, a frameshift mutation in the paired domain of the human protein was
observed in affected family members with oligodontia (Stockton et al., 2000). Homozygous Pax9 mutant mice fail to develop derivatives of the third and fourth pharyngeal pouches and have craniofacial and limb anomalies, as well as teeth that do not progress past the bud stage (Peters et al., 1998).

1.2.2.2 Group III

PAX3 and PAX7 are both essential for the development of neural crest cells, which differentiate into neural structures and melanoblasts (Lang et al., 2007). Pax3 is expressed in the dorsal neural tube as well as in a migratory population of neural crest cells that originates from this area (Goulding et al., 1991). These neural crest cells differentiate into a variety of cell types including melanocytes, peripheral and enteric ganglia, Schwann cells, and craniofacial bone and cartilage. They also contribute to the formation of a variety of organs including the adrenal gland, thymus, parathyroid, and heart (Chi and Epstein, 2002). Pax3 is downregulated in neural crest-derived cells later on in development (Goulding et al., 1991), with the exception of melanocytes (Galibert et al., 1999).

PAX3 and PAX7 are also expressed throughout the somites with their expression eventually restricted to cells that will develop into skeletal muscle (Lang et al., 2007). Once these cells start to differentiate and express myogenic markers, the expression of PAX3 and PAX7 is downregulated (Goulding et al., 1994; Williams and Ordahl, 1994). The two genes are also expressed in adult muscle satellite cells, which initiate regeneration following injury (Seale et al., 2000; Buckingham et al., 2003). Pax7 is particularly important in this cell population as it acts as an anti-apoptotic factor and promotes a quiescent state, which cannot be compensated for by Pax3 (Olguin and Olwin, 2004; Relaix et al., 2006).

Mutations in Pax3 have been identified in the Splotch (Sp) mouse (Epstein et al., 1991; Epstein et al., 1993). Homozygous mutant mice have defects associated with both neural crest and neural tube and die in utero. Along with cardiac abnormalities, including congenital heart disease and persistent truncus
arteriosus, the mice also lack or have a deficient enteric nervous system as well as peripheral ganglia (Lang et al., 2000). The Splotch mouse is similar to human neural crest-related disorders such as DiGeorge syndrome in that they both show similar congenital heart diseases. Pax7 expression is upregulated and its expression area is expanded in Splotch mice thereby allowing it to partially compensate for the lack of Pax3 expression in the neural tube and somites (Borycki et al., 1999). PAX3 mutations are also found in individuals with Waardenburg syndrome, which is an autosomal dominant disorder characterized by deafness and defects in pigmentation (Baldwin et al., 1992; Tassabehji et al., 1992). The deafness observed is thought to be a result of defects in the neural crest cells that are found in the middle ear (Chi and Epstein, 2002). While heterozygous Splotch mice do not appear to have any auditory defects (Steel and Smith, 1992), they do show a pigmentation defect in the form of a white spot on their bellies analogous to the white forelock seen in Waardenburg patients. Hypaxial muscle derivatives are absent in homozygous Splotch embryos and limb muscle defects are also seen in some individuals with Waardenburg syndrome and PAX3 mutations (Hoth et al., 1993). Mice deficient for Pax7 lack satellite cells, which suggests that this gene is required for the specification and/or maintenance of this muscle cell lineage (Seale et al., 2000). Homozygous mutant mice lacking Pax7 develop skeletal muscle, albeit deficiently, and die shortly after birth possibly due to a lack of satellite cells resulting in the inability of the skeletal muscle to regenerate (Seale et al., 2000).

1.2.2.3 Group IV

PAX4 and PAX6 are both expressed in the embryonic pancreatic bud (Lang et al., 2007). As the pancreas grows and differentiates, the widespread expression of PAX4 and PAX6 becomes limited to the islet cell precursors (Lang et al., 2007). Pax4 is restricted to the insulin-producing β cells (Sosa-Pineda et al., 1997), whereas Pax6 is expressed by all four endocrine cell types (St-Onge et al., 1997). Pax6 is also expressed by the developing central nervous system, including the neural tube, pituitary gland, olfactory bulbs, and olfactory epithelium of the nose (Walther and Gruss, 1991). Pax6 is also considered to be the master regulator of eye development, with expression occurring at each
stage of induction of the main components. *Pax6* expression is initially observed in the optic sulcus, which goes on to form the optic vesicle, and then the optic cup (Walther and Gruss, 1991). It is also expressed in the optic stalk, the lens, the cornea, and both layers of the neural retina, (Walther and Gruss, 1991). *Pax6* is also necessary for the maintenance of the pluripotent phenotype in retinal progenitor cells (RPCs), because if it is inactivated immediately following the formation of the RPCs, their fate is restricted to develop only amacrine cells and none of the other five retinal neuron cell types (Marquardt et al., 2001).

In mice deficient in *Pax6*, the pancreas lacks the glucagon-producing α cells, and while the β, δ, and γ cells are present, they do not form properly organized islets (St-Onge et al., 1997). In contrast, a lack of *Pax4* results in the loss of the insulin-producing β cells and the somatostatin-producing δ cells (Sosa-Pineda et al., 1997). This lack of insulin-producing cells is likely the reason *PAX4* mutations have been associated with type II diabetes (Shimajiri et al., 2001; Mauvais-Jarvis et al., 2004). In terms of eye development, *PAX6* mutations in humans lead to aniridia, colobomas, cataracts, and blindness (Jordan et al., 1992; Glaser et al., 1994). Mutations in murine *Pax6* lead to small eye (sey) (Hill et al., 1991), with homozygous mutant mice lacking eyes and failing to develop nasal cavities.

### 1.2.2.4 Group II

*Pax2*, *Pax5* and *Pax8* are expressed in a variety of tissues, both together and separately. All three family members are expressed at the midbrain-hindbrain boundary (Nornes et al., 1990; Plachov et al., 1990; Asano and Gruss, 1992; Stoykova and Gruss, 1994; Joyner, 1996), while *Pax2* and *Pax8* are co-expressed during kidney development (Dressler et al., 1990; Plachov et al., 1990). Individually, *Pax2* is expressed in the eye and inner ear (Nornes et al., 1990; Torres et al., 1996), *Pax8* in the thyroid (Plachov et al., 1990), and *Pax5* is essential for B-cell differentiation (Urbanek et al., 1994).
**PAX2**

*Pax2* expression is observed in the developing mouse midbrain-hindbrain boundary, spinal cord, eyes and ears (Nornes et al., 1990) as well as the kidneys (Dressler et al., 1990). In the central nervous system, *Pax2* is expressed in the neural tube and subsequently along the length of both the rhombencephalon and spinal cord (Nornes et al., 1990). During eye development, it is first observed when the optic vesicle comes in contact with the surface ectoderm to form the optic cup (Nornes et al., 1990). Later on *Pax2* expression is also found in the optic disc and optic nerve (Nornes et al., 1990). In the ear, expression is limited to the areas of the otic vesicle that will form the neuronal portions of the inner ear: the sacculus and the cochlea (Nornes et al., 1990). *Pax2* is expressed throughout kidney development beginning in the pronephric tubules and the extending nephric duct (Dressler et al., 1990). Expression continues in the mesonephros, ureter, mesenchymal condensations and collecting ducts (Dressler et al., 1990). During later stages of development expression diminishes but persists near the perimeter of the growing kidney (Dressler et al., 1990). This coincides with the formation of the S-shaped bodies prior to terminal differentiation (Dressler and Douglass, 1992). PAX2 has also been shown to play a role in the survival of the collecting ducts by acting as an anti-apoptotic factor (Torban et al., 2000).

Mice lacking *Pax2* fail to develop kidneys, ureters, and genital organs (Torres et al., 1995) and show defective development of the eye and inner ear (Torres et al., 1996). *Pax2* expression has been demonstrated in the hyperproliferative cystic tissue of the kidneys of the *cpk* mouse, a model of polycystic kidney disease (Ostrom et al., 2000). In humans, *PAX2* mutations are associated with the autosomal dominant renal-coloboma syndrome, which is characterized by both renal and ocular defects (Eccles, 1998).

**PAX5**

*Pax5* plays an important role in the development of the midbrain as well as in B-cell differentiation (Urbanek et al., 1994). During embryogenesis, *Pax5* is expressed in the developing central nervous system in both the brain and neural tube. Strong expression is observed in the posterior region of the
mesencephalon as well as on both ventral sides of the third ventricle (Adams et al., 1992). In the neural tube, *Pax5* is expressed in the border region that separates the mitotically active neuronal stem cells of the ventricular zone from the differentiating postmitotic neurons of the mantle layer (Adams et al., 1992). Unlike the other two group members, *Pax2* and *Pax8*, *Pax5* is not expressed in the rhombencephalon (Nornes et al., 1990; Plachov et al., 1990; Adams et al., 1992). Later on in development, expression of *Pax5* shifts to the fetal liver where its increase coincides with the onset of B lymphopoiesis (Rolink and Melchers, 1991; Adams et al., 1992). In the adult mouse, *Pax5* is expressed by the B lymphocytes during the early stages of B-cell differentiation, from the early pro-B-cell through to the mature B-cell (Adams et al., 1992). It plays an important role in the commitment of the pluripotent pro-B cells to the B-cell lineage by blocking their ability to differentiate into other lineages (Nutt et al., 1999). *Pax5* expression has also been found in the adult mouse testis (Adams et al., 1992).

B-cell development in *Pax5* knockout mice is blocked at the pro-B-cell stage (Urbanek et al., 1994). In the absence of *Pax5*, pro-B-cells are uncommitted and differentiate into other cell types including natural killer cells and macrophages (Nutt et al., 1999).

**PAX8**

*PAX8* is expressed in the developing central nervous system, inner ear, placenta, kidneys, and thyroid. Akin to *Pax2*, *Pax8* is expressed transiently along the length of the neural tube (Plachov et al., 1990). In the developing brain, *Pax8* expression is located in the hindbrain, with strong expression occurring in the myelencephalon and weaker expression in the metencephalon (Nornes et al., 1990; Plachov et al., 1990; Asano and Gruss, 1992).

*Pax8* is one of the first genes expressed in the developing otic placode of several vertebrates including fish, frogs, and mice (Pfeffer et al., 1998; Mackereth et al., 2005) and there is evidence that the entire inner ear originates from cells expressing *Pax8* (Grote et al., 2006). Despite the seeming
importance of \textit{Pax8} in the developing ear, its early expression is quickly replaced by \textit{Pax2} (Bouchard et al., 2010).

\textit{PAX8} mRNA has been found in both murine (Kozmik et al., 1993) and human placental cells, with levels comparable to those found in the thyroid (Ferretti et al., 2005). Expression is mainly found in the trophoblast cells, along with sodium/iodide symporter (\textit{NIS}) and the iodothyronine deiodinases, D2 and D3, which are required for supplying the iodide necessary for the synthesis of fetal thyroid hormones (Bidart et al., 2000; Chan et al., 2003; Huang et al., 2003; Ferretti et al., 2005).

During kidney development \textit{Pax8} expression often overlaps with that of \textit{Pax2}, with \textit{Pax8} being expressed primarily by the induced tissues (mesenchymal condensations and S-shaped bodies) (Plachov et al., 1990). \textit{PAX8} is found in the induced mesenchyme surrounding the ureteric bud (Eccles et al., 1995), with high expression occurring at the time the mesenchyme folds into comma- and S-shaped bodies (Poleev et al., 1992). Once differentiation begins to occur, transcription of \textit{PAX8} is downregulated (Eccles et al., 1995).

\textit{PAX8} expression in the thyroid is first observed at the site of evagination from the floor of the pharynx and continues throughout development and into the adult gland (Plachov et al., 1990; Zannini et al., 1992). Expression is necessary for proper differentiation of endoderm cells into thyroid follicular cells (TFCs), which synthesize the thyroid hormones (Mansouri et al., 1998; Lang et al., 2007). The synthesis of these thyroid hormones requires expression of thyroglobulin (\textit{Tg}), thyroperoxidase (\textit{TPO}), and sodium/iodide symporter (\textit{NIS}), all thyroid-specific genes regulated by \textit{PAX8} (Pasca di Magliano et al., 2000).

No obvious ear defects have been observed in mice lacking \textit{Pax8} (Mansouri et al., 1998), suggesting that \textit{Pax2} is able to compensate (Bouchard et al., 2010). While \textit{Pax8} heterozygous mutant mice are healthy and fertile, the homozygous mutant mice do not grow properly and have smaller thyroids that lack TFCs (Mansouriri et al., 1998). Unless treated with thyroid hormones, these mice will
die around the time of weaning (Mansouri et al., 1998; De Felice and Di Lauro, 2004). The thyroid is the only Pax8-expressing organ that is affected, most likely due to compensation by Pax2 and/or Pax5 in the kidney and central nervous system (Dressler et al., 1990; Asano and Gruss, 1992; Urbanek et al., 1994; Mansouri et al., 1998).

Unlike the Pax8 heterozygous mutant mice, which do not show an obvious defect in the thyroid (Mansouri et al., 1998), humans missing a copy of PAX8 have severely affected thyroids (Macchia et al., 1998). PAX8 mutations have been associated with congenital hypothyroidism, which is characterized by a lack of thyroid hormones (Macchia et al., 1998; Congdon et al., 2001). The reduction in size of the gland suggests that PAX8 is involved in the proliferation or maintenance of the differentiated thyroid cell populations (Macchia et al., 1998).

1.2.3 PAX Regulation of Transcription

As transcription factors the PAX proteins regulate a variety of genes, frequently working in co-operation with other proteins. Along with having similar expression patterns, members of the same subgroup often recognize similar binding sites (Chi and Epstein, 2002). As would be expected, the genes regulated by the PAX proteins are often, but not always, related to the tissues in which they are expressed. Once such example involves the Group I proteins Pax1 and Pax9, which have been shown to activate transcription of the Bapx1 gene via direct interaction with its promoter (Rodrigo et al., 2003). Akin to Pax1 and Pax9, Bapx1 is expressed in the sclerotome (Tribioli et al., 1997).

1.2.3.1 GROUP III

Pax3, which is expressed in neural crest cells, regulates other genes associated with their development. Together with Sox10, Pax3 activates an enhancer in the c-RET gene, which is required for enteric ganglia formation (Lang et al., 2000). Pax3 and Sox10 also interact in the activation of MITF, which is
required for melanocyte development (Bondurand et al., 2000; Potterf et al., 2000; Verastegui et al., 2000). Pax3 then competes with Mitf for access to an enhancer region in the Dct gene where, upon binding, it represses transcription thereby blocking terminal differentiation of melanoblasts (Lang et al., 2005). Dct and TRP-1, another gene regulated by Pax3 (Galibert et al., 1999), both encode enzymes involved in melanin synthesis. Pax3 also represses Msx2, which is involved in cardiac neural crest development, by binding directly to its promoter (Kwang et al., 2002). During limb muscle development Pax3 regulates the expression of c-met, a tyrosine kinase receptor required for myoblast migration (Epstein et al., 1996). PAX3 is also able to activate the Bcl-x gene promoter, thereby regulating transcription of the anti-apoptotic protein Bcl-XL (Margue et al., 2000). Pax3 is able to act as a repressor by interacting with corepressors, such as HIRA (Magnaghi et al., 1998) and hDaxx (Hollenbach et al., 1999), through its homeodomain and octapeptide domain.

1.2.3.2 GROUP IV

Similar to Pax3 and Sox10, Pax6 and Sox2 form a complex and cooperatively activate the enhancer elements of lens-specific crystallin genes (Kamachi et al., 2001). Pax6 along with Pax2, both of which are expressed in the developing eye, transactivate Mitf (Baumer et al., 2003). Mitf encodes a transcription factor that plays a role in the development and differentiation of the retinal pigment epithelium, akin to its role in melanocyte development. Pax6 has also been shown to regulate the expression of Pdx1, a gene essential in pancreas development and the maintenance of the insulin-producing $\beta$ cells (Samaras et al., 2002). There is also evidence that both Pax6 and Pax4 are able to directly regulate transcription of the pancreatic hormones somatostatin, glucagon, and insulin, although they appear to do so in different capacities with Pax6 acting as a transactivator (Sander et al., 1997) and Pax4 as a repressor (Smith et al., 1999).
**1.2.3.3 GROUP II**

PAX2 regulates transcription of *WT1*, which is involved in kidney development (Dehbi et al., 1996; McConnell et al., 1997). It also directly activates expression of the *Gdnf* gene, the product of which is necessary for proper ureteric bud outgrowth leading to kidney development (Brophy et al., 2001). Several B-cell-specific genes that encode cell surface proteins are regulated by Pax5. Both *CD19* and *mb-1* are activated, while *PD-1* is repressed (Nutt et al., 1997; Nutt et al., 1998). Transcriptional regulation of *mb-1* is accomplished with the aid of the Ets family of transcription factors, which are recruited by Pax5 and form a complex bound to the promoter (Fitzsimmons et al., 1996). Pax5 also regulates expression of the transcription factors *N-myc* and *LEF-1* (Nutt et al., 1998). Its role as a repressor is carried out via the interaction of the Groucho family of corepressors, such as Grg-4, with its octapeptide domain (Eberhard et al., 2000).

**PAX8**

During kidney development PAX8 activates transcription of *WT1*, akin to PAX2 (Dehbi and Pelletier, 1996; Fraizer et al., 1997). PAX8 is also able to directly regulate the thyroid-specific genes *thyroglobulin* (*Tg*), *thyroperoxidase* (*TPO*), and the *sodium/iodide symporter* (*NIS*) (Zannini et al., 1992; Ohno et al., 1999). In the case of thyroglobulin transcription, PAX8 and thyroid transcription factor-1 (TTF-1) have been shown to co-operate (Pasca di Magliano et al., 2000; Espinoza et al., 2001) via a direct interaction (Di Palma et al., 2003). The co-activator TAZ is able to further enhance transcriptional activation by Pax8 and TTF-1 by physically associating with them (Di Palma et al., 2009). On the *TPO* gene promoter, the p300 protein acts as a Pax8 co-activator (De Leo et al., 2000). The retinoblastoma protein, Rb, has been identified as a Pax8 co-activator of both *TPO* and *NIS* gene transcription (Miccadei et al., 2005). PAX8 has been shown to synergistically activate transcription of *NIS* with the transcription factor p65 (Nicola et al., 2010).

Outside of their individual tissue-specific gene regulation, PAX2, PAX5, and PAX8 have been reported to repress transcription of the *p53* gene, whose product is involved in cell cycle regulation and is often mutated in cancer
(Stuart et al., 1995). In gliomas, PAX8 activates transcription of two genes involved in the main catalytic holoenzyme of telomerase: human telomerase reverse transcriptase (hTERT) and human telomerase RNA (hTR) (Chen et al., 2008). PAX8 has also been shown to regulate transcription of the bel-2 proto-oncogene (Hewitt et al., 1997).

1.2.4 PAX and Cancer

In addition to regulating the transcription of certain genes that are known to be involved in cancer, several members of the PAX family are expressed in a variety of cancers themselves (Muratovska et al., 2003; Robson et al., 2006; Lang et al., 2007). As with the genes they regulate, their expression in cancerous cells is often tissue-specific and related to expression patterns observed during development (Lang et al., 2007). Tumour-specific mutations to the genes give them an oncogenic role while persistent expression often characterizes tumours. In some cases, treatment with antisense oligonucleotides or small interfering RNA (siRNA) targeting the specific PAX gene being expressed reduces cell proliferation and tumorigenesis (Gnarra and Dressler, 1995; Bernasconi et al., 1996; Scholl et al., 2001; Muratovska et al., 2003).

1.2.4.1 CHROMOSOMIC TRANSLOCATIONS

Several members of the PAX family are involved in chromosomal translocations that have been observed in various cancers (Barr, 1997). A fusion protein made up of the DNA-binding domain of PAX3 and the transcriptional activation domain of a forkhead family gene product, FKHR, can result in pediatric alveolar rhabdomyosarcoma (Barr et al., 1993; Galili et al., 1993; Shapiro et al., 1993). A similar translocation involving PAX7 has also been reported (Davis et al., 1994). The PAX3/FKHR fusion protein is a more effective transcription factor than wildtype PAX3 (Fredericks et al., 1995). Treatment with antisense oligonucleotides targeting the PAX3/FKHR fusion protein results in the induction of apoptosis (Bernasconi et al., 1996). PAX3/FKHR, as well as aberrantly expressed PAX3, is able to regulate transcription of the anti-
apoptotic protein Bcl-X_L, suggesting a potential mechanism by which rhabdomyosarcoma tumours are able to evade apoptosis (Margue et al., 2000). Overexpression of PAX3, PAX3/FKHR, or Bcl-X_L can also rescue tumour cells from apoptosis resulting from treatment with antisense oligonucleotides (Margue et al., 2000). A chromosomal translocation combining the powerful Eµ enhancer of the IgH gene with PAX5 has been described in both large-cell and lymphoplasmacytoid lymphomas (Busslinger et al., 1996; Iida et al., 1996).

**PAX8**
The chromosomal translocation t(2;3)(q13;p25) yields a fusion protein consisting of the PAX8 DNA-binding paired domain and homeodomain joined to peroxisome proliferator-activated receptor (PPAR) gamma 1 (Poleev et al., 1992; Kozmik et al., 1993). Expressed in both thyroid follicular carcinomas and adenomas (Kroll et al., 2000; Cheung et al., 2003), PAX8-PPARγ1 is able to inhibit transcriptional activation by wild-type PPARγ and may also deregulate PAX8 pathways (Kroll et al., 2000). It is believed that the fusion of PAX8 and PPARγ1 creates a protein that plays an important role in thyroid cancer development (Kroll et al., 2000).

### 1.2.4.2 PERSISTENT EXPRESSION
Pax2 is expressed in Wilms’ tumour (Dressler and Douglass, 1992; Eccles et al., 1992) and has also been observed in all of the major histological types of renal cell carcinomas (RCCs) with the exception of urothelial RCCs (Daniel et al., 2001). When treated with antisense oligonucleotides, PAX2 expression is downregulated and growth slows in human RCCs (Gnarra and Dressler, 1995). Bladder and ovarian cancer cell lines also undergo apoptosis following treatment with PAX2 siRNAs (Muratovska et al., 2003).

Expression of PAX3 and PAX7 has been found in Ewing’s sarcoma and embryonal rhabdomyosarcoma, and PAX3 is expressed in melanoma cell lines (Barr et al., 1999). In embryonal rhabdomyosarcoma cells, cell death is triggered following treatment with antisense oligonucleotides targeting wild-
type \textit{PAX3} and \textit{PAX7} as well as the \textit{PAX3/FKHR} fusion product (Berasconi et al., 1996). \textit{PAX3} has been shown to be necessary for the survival of melanoma cell lines (Muratovska et al., 2003) with \textit{PAX3}-expressing melanoma cell lines undergoing apoptosis following treatment with antisense oligonucleotides (Scholl et al., 2001).

\textbf{PAX8}

\textit{PAX8} is expressed in renal collecting duct carcinoma (Albadine et al., 2010), nephrogenic adenoma and clear cell carcinoma of the lower urinary tract (Tong et al., 2008), Wilms’ tumour (Poleev et al., 1992), bladder cancer (Pellizziari et al., 2006), ovarian cancer (Bowen et al., 2007), and glioma (Chen et al., 2008).

\subsection{1.3 The E2F Gene Family}

The E2F gene family encodes a group of transcription factors that are involved in the regulation of cell cycle progression. The eight members in the E2F family are divided into two subgroups based on similarities in structure and function (Figure 1.2). E2F1, E2F2, and E2F3 are transcriptional activators, while E2F4-8 function as repressors (DeGregori and Johnson, 2006; Chong et al., 2009). The E2F proteins, with the exception of E2F7 and E2F8, function by forming a heterodimeric complex with one of the E2F dimerization proteins (DP1, DP2/3, or DP4) (DeGregori and Johnson, 2006; Stanelle and Putzer, 2006). The activity of the E2F transcription factors is regulated by the retinoblastoma protein (Rb) and the related pocket proteins, p130 and p107. These proteins are able to bind to the E2F proteins and block access to the transactivation domain, thereby resulting in the restriction of cell cycle progression (Stanelle and Putzer, 2006). The E2F proteins vary in which repressor protein they bind to. E2F1-3 can only be bound by Rb when it is in the hypophosphorylated state (Lees et al., 1993). E2F4 can associate with all of the pocket proteins, whereas E2F5 is bound primarily by p130 (Trimarchi and Lees, 2002; Attwooll et al., 2004; Dimova and Dyson, 2005). E2F6-8 do not associate with any of the repressor proteins due to a lack of a transactivation
domain (Stanelle and Putzer, 2006). Along with the transactivation and
dimerization domains, the E2F proteins contain other structural domains
including a DNA-binding domain and domains required for association with
other proteins, including the Rb-family proteins.

Figure 1-2: The E2F Transcription Factor Family
The E2F family is divided into activators (E2F1–E2F3) and repressors (E2F4–E2F8), which
bind to DNA via their DNA-binding domain(s) (DBD). E2F1–E2F6 bind DNA as heterodimers
with one of three dimerization partner (DP) proteins, whose dimerization is mediated by the
leucine zipper (LZ) and marked box (MB) domains. Rb binds within the transactivation domain
(Rb) of E2F1–E2F3. The nuclear localization signal (NLS) sequence, which is adjacent to the
cyclin A-binding site (CycA), localizes the activators to the nucleus. E2F4 and E2F5 have
nuclear export signals (NES) that initiate their export to the cytoplasm. Adapted from (Chen et
al., 2009).

1.3.1 E2F Gene Expression During the Cell Cycle
The various members of the E2F transcription factor family are expressed at
different times during the cell cycle (Figure 1.3). During the G0 phase, E2F4
and E2F5 are expressed and, in combination with the pocket proteins and other
co-repressors, act to repress the transcription of E2F target genes involved in
cell cycle entry (Chen et al., 2009). Upon stimulation by the necessary signals, a cascade is initiated that results in the phosphorylation of Rb, thereby blocking its function and releasing the E2F repressors (Harbour et al., 1999). This then leads to the synthesis of the activator E2Fs – E2F1, E2F2, and E2F3 – which accumulate in late G1 and activate the expression of genes required to complete the G1/S transition. Once DNA synthesis is completed in S phase, transcription is attenuated in G2 by the other E2F repressors, E2F6-8 (Chen et al., 2009).

Figure 1-3: E2F Expression During the Cell Cycle
During G0, E2F4 and E2F5 interact with the pocket proteins and repress transcription of E2F target genes, indicated by the purple line labeled E2F target, that promote entry into G1. Phosphorylation of Rb releases the repressors and leads to the synthesis of the activator E2Fs, E2F1-3, in late G1. This results in the transcription of genes that drive cells into S phase, where DNA replication occurs (green line). E2F6-8 then repress DNA transcription as cells move from S phase to G2. Adapted from (Chen et al., 2009).

1.3.2 E2F Genes and Cancer
Genetic alterations resulting in a loss of functional Rb are believed to cause cancer by leading to increased levels of free E2F that drive cell cycle progression and uncontrolled cell proliferation. Given the classification of E2F proteins as either activators or repressors, it might be expected that the activators would act as oncogenes, whereas the repressors would act as tumour
suppressors (Chen et al., 2009). In reality, however, all of the E2Fs appear to have a role in promoting tumour progression. In addition to increased expression, other genetic alterations have also been observed, including: amplification, deletion, mutation, and decreased expression. The E2F family member that is probably most often altered in cancer is E2F1 (Table 1.1).

### 1.3.3 E2F1 Expression

As can be seen in Table 1.1, the most common alteration to E2F1 in cancer is an increase in expression. In normal cells, *E2F1* gene expression is regulated by the cell cycle (Johnson et al., 1994). The activator E2Fs, E2F1-3, activate transcription of *E2F1* (Hsiao et al., 1994; Johnson et al., 1994; Neuman et al., 1994; Araki et al., 2003) while E2F4 acts to repress it (Araki et al., 2003). E2F1 autoregulation of its own promoter shows the greatest increase in activity (Hsiao et al., 1994). Outside of the E2F family, Myc has also been shown to regulate *E2F1* gene transcription by interacting directly with the promoter and playing a role in loading E2F1 onto it (Fernandez et al., 2003; Leung et al., 2008).

E2F1 is key to the control of the G1/S transition and thus its expression needs to be properly regulated. Increased transcription of the *E2F1* gene could lead to aberrant activation of its target genes followed by inappropriate driving of the cell cycle from G1 to S phase. If the cell cycle is able to proceed unchecked, this could result in uncontrolled cell growth and proliferation resulting in tumour formation.

It has been proposed that a basal level of E2F1 may be maintained in cells in order to ensure that the cells are able to respond to signals stimulating proliferation prior to entry into the S phase (Li et al., 2011). This could be important in cancer cells in which there is an alteration to a member of the Rb-E2F1 pathway as this would result in the release of free E2F1 that could drive the cell cycle forward.
# Table 1-1: Genetic alterations of E2F1 in human cancers

<table>
<thead>
<tr>
<th>Cancer</th>
<th>Amplification</th>
<th>Increased Expression</th>
<th>Decreased Expression</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder</td>
<td></td>
<td></td>
<td>☭</td>
<td>Rabbani et al., 1999</td>
</tr>
<tr>
<td>Breast</td>
<td></td>
<td>☭</td>
<td></td>
<td>Han et al., 2003</td>
</tr>
<tr>
<td>Cervical</td>
<td>☭</td>
<td></td>
<td></td>
<td>Wilting et al., 2006; Huang et al., 2007; Wilting et al., 2008</td>
</tr>
<tr>
<td>Colon</td>
<td></td>
<td></td>
<td>☭</td>
<td>Bramis et al., 2004</td>
</tr>
<tr>
<td>Esophageal SCC</td>
<td>☭</td>
<td>☭</td>
<td></td>
<td>Fujita et al., 2003; Ebihara et al., 2004; Mega et al., 2005</td>
</tr>
<tr>
<td>Gastric adenocarcinoma</td>
<td></td>
<td></td>
<td>☭</td>
<td>Lee et al., 2008</td>
</tr>
<tr>
<td>GI stromal</td>
<td></td>
<td>☭</td>
<td></td>
<td>Haller et al., 2005</td>
</tr>
<tr>
<td>Glioblastoma</td>
<td></td>
<td>☭</td>
<td></td>
<td>Alonso et al., 2005</td>
</tr>
<tr>
<td>Hepatocellular carcinoma</td>
<td>☭</td>
<td>☭</td>
<td></td>
<td>Zondervan et al., 2000; Xu et al., 2001; Midorikawa et al., 2004; Midorikawa et al., 2007; Llovet and Bruix, 2008</td>
</tr>
<tr>
<td>Non-small-cell lung</td>
<td>☭</td>
<td>☭</td>
<td></td>
<td>Gorgoulis et al., 2002; Imai et al., 2004; Karakaidos et al., 2004; Tonon et al., 2005; Huang et al. 2007</td>
</tr>
<tr>
<td>Oral SCC</td>
<td></td>
<td></td>
<td>☭</td>
<td>Kwong et al., 2003</td>
</tr>
<tr>
<td>Ovarian</td>
<td></td>
<td>☭</td>
<td></td>
<td>Reimer et al., 2006; Reimer et al., 2007; De Meyer et al., 2009</td>
</tr>
<tr>
<td>Pancreatic ductal</td>
<td></td>
<td>☭</td>
<td></td>
<td>Yamazaki et al., 2003</td>
</tr>
<tr>
<td>Small-cell lung</td>
<td></td>
<td>☭</td>
<td></td>
<td>Eymin et al., 2001; Mega et al., 2005</td>
</tr>
</tbody>
</table>

SCC = squamous cell carcinoma; GI = gastrointestinal. Adapted from (Chen et al., 2009).
1.4 **Background**

The analysis of several human cancer cell lines for the presence of \( PAX \) transcripts and proteins revealed that \( PAX8 \) is frequently expressed (Li et al., 2011), consistent with previous studies (Muratovska et al., 2003; Robson et al., 2006). A subsequent microarray analysis of gene expression changes following PAX8 silencing in A498 cells identified \( E2F1 \), as well as some of its targets, as being downregulated (Li et al., 2011 and unpublished). Following validation using quantitative PCR (qPCR), changes in protein levels were analyzed. It was discovered that both E2F1 and its targets, cyclin A and CDC6, were present in reduced quantities following treatment with siRNA targeting \( PAX8 \). Based on these results, it was hypothesized that PAX8 may directly activate the \( E2F1 \) promoter.

1.5 **Aim**

The overall aim of this study is to characterize the relationship between PAX8 and \( E2F1 \) in cancer cells. Each of the three chapters of results (Ch. 3-5) addresses a specific aspect of this aim:

*Chapter 3* – to determine if PAX8 expression directly affects \( E2F1 \) promoter activity

*Chapter 4* – to establish if PAX8 binds directly to the \( E2F1 \) promoter and, if so, where

*Chapter 5* – to validate the functionality of each predicted PAX8 binding site
CHAPTER 2
– Material and Methods –

A complete list of materials and suppliers can be found in Appendix A. The compositions of buffers and other solutions used are outlined in Appendix B.

2.1 Cell Lines

The cell lines used are listed in Table 2-1.

2.2 Cell Culture

2.2.1 Maintenance

Cells were cultured in DMEM (Invitrogen, USA) supplemented with 10% Fetal Bovine Serum (Bio International Ltd, NZ) and maintained in a 37°C incubator with 5% CO₂. When cells reached a confluence of 80% or more, they were subcultured. The media was aspirated from the flask, 0.25% trypsin-EDTA (Invitrogen, USA) was added, and the flask was returned to the incubator for a few minutes to assist the progress of the enzymatic reaction. Once the cells had lifted off the flask, fresh media was added to neutralize the trypsin and wash the cells off the flask. The cell suspension was then transferred to a tube and centrifuged for 4 minutes at 200 x g. The media was aspirated and the cell pellet was resuspended in fresh media. An appropriate amount of cells was then returned to the flask and topped up to the desired culture volume with fresh media.

2.2.2 Transient Transfection

The plasmids used for transient DNA transfections are listed in Table 2-2, while the small interfering RNAs (siRNAs) used are listed in Table 2-3.
<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Species</th>
<th>Organ</th>
<th>Disease</th>
<th>Original Source</th>
<th>Experimental Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A498</td>
<td><em>Homo sapiens</em></td>
<td>kidney</td>
<td>carcinoma</td>
<td>ATCC</td>
<td>Lab stock</td>
<td>(Giard et al., 1973)</td>
</tr>
<tr>
<td>HEK293</td>
<td><em>Homo sapiens</em></td>
<td>embryonic</td>
<td>---</td>
<td>ATCC</td>
<td>Lab stock</td>
<td>(Shaw et al., 2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kidney</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1</td>
<td><em>Homo sapiens</em></td>
<td>thyroid</td>
<td>carcinoma</td>
<td>Prof. D. Wynford-Thomas (University of Leicester, UK)</td>
<td>Prof. Antony Braithwaite (University of Otago, NZ)</td>
<td>(Challeton et al., 1997)</td>
</tr>
</tbody>
</table>
**Table 2-2: Plasmid DNA used in this study**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Source</th>
<th>Origin</th>
<th>Insert Size</th>
<th>Location</th>
<th>Vector</th>
<th>Reference (if applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCR3.1</td>
<td>Lab stock (Invitrogen, USA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pGL2-basic</td>
<td>Lab stock (Promega, USA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pdx11-LacZ</td>
<td>Dr. Phil Daniel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(pCMV-β-gal)</td>
<td>(University of Otago, NZ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCMV5-Pax8</td>
<td>Prof. Robert Di Lauro (University of Naples, Italy)</td>
<td>mouse Pax8 cDNA</td>
<td>2.3 kb</td>
<td>ClaI/BamHI</td>
<td>pCMV5</td>
<td>(Zannini et al., 1992)</td>
</tr>
<tr>
<td>pCMV-E2F1</td>
<td>Assoc. Prof. Stephen Hsu (University of Florida, USA)</td>
<td>human E2F1 cDNA</td>
<td>2.7 kb</td>
<td>BamHI</td>
<td>pCMVneoBam</td>
<td>(Helin et al., 1993)</td>
</tr>
<tr>
<td>pE2F1(-242)</td>
<td>Assoc. Prof. Masa-Aki Ikeda (Tokyo Medical and Dental University, Japan)</td>
<td>human E2F1 promoter</td>
<td>0.3 kb</td>
<td>SmaI/HindIII</td>
<td>pGL2-basic</td>
<td>(Johnson et al., 1994)</td>
</tr>
<tr>
<td>pE2F1(-728)*</td>
<td>Prof. Kiyoshi Ohtani (Kwansei Gakuin University, Japan)</td>
<td>human E2F1 promoter</td>
<td>0.8 kb</td>
<td>SacI/HindIII</td>
<td>pGL2-basic</td>
<td>(Johnson et al., 1994)</td>
</tr>
</tbody>
</table>

* Prof. Ohtani provided the following information regarding pE2F1(-728): “during expansion in DH5α, deletion occurs in a portion of the plasmid. The deleted mutant seems to be inactive in terms of luciferase expression and contamination of the deleted mutant does not hurt reporter assay.”
Table 2-3: siRNAs used in this study

<table>
<thead>
<tr>
<th>siRNA target (target)</th>
<th>Name (abbreviation)</th>
<th>Sequence (5' =&gt;3')</th>
<th>Manufacturer (siRNA ID)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-target</td>
<td>ON-TARGETplus</td>
<td>N/A</td>
<td>Dharmacon, USA</td>
</tr>
<tr>
<td></td>
<td>Non-targeting Pool</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(siControl)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAX8</td>
<td>ON-TARGETplus</td>
<td>CAUCCG GCCUGGAGUUA</td>
<td>Dharmacon, USA</td>
</tr>
<tr>
<td></td>
<td>SMARTpool</td>
<td>CCUCACAA CUCAUCAUGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PAX8 (siPAX8)</td>
<td>CGACAUCUCUCGCCACGC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAGCAUAGCUGGCCAGUA</td>
<td></td>
</tr>
</tbody>
</table>

2.2.2.1 DNA Transfection

HEK293 cells were seeded at a concentration of 60,000 cells per well into the required number of wells of a 24-well plate 24 hours prior to DNA transfection. This was done in order to obtain a confluence of 80-90% at the time assays were to be carried out. FuGENE 6 Transfection Reagent (Roche Applied Science, Germany) was used for the transfection of the plasmid DNA according to the manufacturer’s instructions. A 6:1 Fugene:DNA ratio was used for transfection of a total of 600 ng of DNA, consisting of 75 ng of the selected luciferase reporter construct, 25 ng of pCMV-β-gal, with the remaining 500 ng made up of a combination of pCMV5-Pax8, pCR3.1, and/or pCMV-E2F1. The pCR3.1 plasmid was included in order to keep the amount of DNA the same between all of the samples.

2.2.2.2 siRNA Transfection

The siRNA for transfection was prepared using OPTI-MEM and Lipofectamine RNAiMAX (both Invitrogen, USA) according to the manufacturer’s instructions. During the incubation of the transfection mixture, K1 cells were harvested and prepared for seeding at a concentration of 60,000 cells per well for the siControl samples and 80,000 cells per well for the siPAX8 samples in a
24-well plate. This would result in a confluence of 80-90% at the time of harvesting for use in subsequent assays. After the incubation, 100 µL of transfection mix was added to each well followed by a 400 µL aliquot of K1 cell suspension drop-wise. The plate was tapped equally on each side to ensure that the cells were evenly distributed in a monolayer and then placed in the incubator.

### 2.2.2.3 siRNA and Plasmid DNA Co-transfection

For co-transfection of both siRNA and plasmid DNA, cells were transfected with siRNA first, as described in Section 2.2.2.2. Transfection of plasmid DNA was then carried out 24 hours later, as outlined in Section 2.2.2.1.

### 2.2.3 Reporter Gene Analysis

#### 2.2.3.1 Harvesting Cell Lysate

Cells were harvested 48 hours post-DNA transfection by aspirating the media and adding 200 µL of 0.25% trypsin-EDTA (Invitrogen, USA) to each well. The cells were flushed off the plate and then transferred to a microfuge tube. The well was then washed with 1 mL of PBS and added to the tube. The cell suspensions were centrifuged for 10 minutes at 300 x g and 4°C. The supernatant was aspirated and the pellet was washed in 500 µL of PBS and again centrifuged at 300 x g and 4°C for 10 minutes. During this time the cell lysis buffer was prepared as a 1:5 dilution of 1X Cell Culture Lysis Reagent (Promega, USA) in dH₂O, with 50 µL per sample plus a blank. The supernatant was aspirated and the pellet was completely resuspended in cell lysis buffer and incubated on ice for 5 minutes. The samples were then centrifuged at 16,100 x g and 4°C for 5 minutes. The resulting supernatant was then divided between two tubes – one containing 10 µL for the β-galactosidase (β-gal) assay and the rest for use in the luciferase assay – and kept on ice.
2.2.3.2 Luciferase Assay

The Luciferase Assay Reagent (Promega, USA) was prepared as per the manufacturer’s instructions and stored in 1 mL aliquots at -80°C. For the assay, the luciferase reagent was removed from -80°C, quickly thawed in a 37°C water bath, and then kept on ice. The assay was performed by gently mixing 30 µL of cell lysate with 80 µL of luciferase reagent directly in the well of an opaque white 96-well plate. Two reactions were carried out simultaneously and the luminescence of each sample was measured well-by-well using the Synergy 2 Multi-Mode Microplate Reader (BioTek, USA) and the accompanying Gen5 software (version 1.01.9).

2.2.3.3 β-galactosidase Assay

The β-gal substrate was prepared at a volume of 100 µL per sample and consisted of a mixture of PBS, ONPG, MgCl₂, and β-mercaptoethanol. The substrate was added to the 10 µL aliquot of cell lysate, mixed, and incubated at 37°C in a heat block until the reaction mixture was bright yellow in colour. A 100 µL aliquot of each sample was transferred to a transparent 96-well plate and the absorbance was measured at 410 nm using the Synergy 2 plate reader.

2.2.3.4 Data Analysis

The results obtained from the luciferase assay were normalized relative to the β-gal activity levels to yield the relative luciferase units (RLU). This was done in order to account for differences in transfection efficiency between the samples. Normalization of the luciferase assay results to the β-gal activity was calculated separately for each promoter-reporter construct. The sample with the highest β-gal activity value within a set was selected as the calibrator. The β-gal values for each of the remaining samples were then divided by the value of the calibrator to yield the β-gal ratio, with that of the calibrator sample having a value of 1.0. The normalized RLU value of each of the samples was then calculated by dividing its luciferase value by its β-gal ratio. Examples using pGL2-basic, pE2F1(-242), and pE2F1(-728) co-transfected with PAX8 are
shown in Table 2-4. The mean of the RLU values obtained for each individual PAX8/plasmid combination were plotted with error bars of +/- 1 SEM.

<table>
<thead>
<tr>
<th>PAX8</th>
<th>Plasmid</th>
<th>Luc Value</th>
<th>β-gal Value</th>
<th>β-gal Ratio</th>
<th>RLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ng</td>
<td>pGL2-basic</td>
<td>350</td>
<td>0.757</td>
<td>0.650</td>
<td>538.639</td>
</tr>
<tr>
<td>400 ng</td>
<td>pGL2-basic</td>
<td>314</td>
<td>1.165*</td>
<td>1.000</td>
<td>314.000</td>
</tr>
<tr>
<td>0 ng</td>
<td>pE2F1(-242)</td>
<td>25801</td>
<td>0.769</td>
<td>0.866</td>
<td>29793.612</td>
</tr>
<tr>
<td>400 ng</td>
<td>pE2F1(-242)</td>
<td>97460</td>
<td>0.888*</td>
<td>1.000</td>
<td>97460.000</td>
</tr>
<tr>
<td>0 ng</td>
<td>pE2F1(-728)</td>
<td>45374</td>
<td>0.475</td>
<td>0.629</td>
<td>72120.779</td>
</tr>
<tr>
<td>400 ng</td>
<td>pE2F1(-728)</td>
<td>171045</td>
<td>0.755*</td>
<td>1.000</td>
<td>171045.000</td>
</tr>
</tbody>
</table>

* indicates the calibrator for each plasmid

Statistical analysis of the luciferase results was performed using GraphPad Prism (version 4.0b). The unpaired two-tailed t test was used to analyze the statistical significance between two samples while the one-way ANOVA and Tukey’s post test were applied to data sets containing three or more samples.

### 2.3 Protein Expression Analysis

#### 2.3.1 Total Protein Isolation

Cells were harvested by trysinization, the well rinsed with PBS, the entire mixture added to a 15 mL Falcon tube, and centrifuged for 5-10 minutes at 250 x g. The supernatant was completely removed and the pellet was gently resuspended in PBS. The cells were then centrifuged for another 10 minutes at 250 x g during which time the lysis buffer was prepared by adding protease inhibitors to RIPA buffer and kept on ice. The supernatant was again aspirated and the pellet was resuspended in the appropriate amount of lysis buffer, which
was dependent on the cell number and confluency at the time of harvesting. The cell lysate was transferred to an ice-cold tube and incubated on ice for 30 minutes. The samples were then centrifuged at 16,100 x g for 20 minutes at 4°C. The supernatant was transferred to a fresh ice-cold tube and stored at -80°C until needed.

2.3.2 Protein Quantification

Protein samples were quantified using the colourimetric BCA Protein Assay Kit (Thermo Scientific, USA) according to the manufacturer’s instructions. In order to determine the protein concentration, a series of BSA standards was used to plot a standard curve.

2.3.2.1 Bovine Serum Albumin Standard Curve

The BSA standards were prepared on ice from a 10 mg/mL BSA stock included with the BCA Protein Assay Kit (Thermo Scientific, USA) as outlined in Table 2-5. The standards were divided into two 20 μL aliquots in a 96-well plate. The BCA buffer was then prepared according to the manufacturer’s instructions. The absorbance of the standards was then measured at 570 nm using the Anthos Reader 2010 microplate reader (Anthos Labtec Instruments, Austria) and the values were used to plot a standard curve.
Table 2-5: BSA standard preparation

<table>
<thead>
<tr>
<th>MQ-H₂O (µL)</th>
<th>Stock BSA (µL)</th>
<th>Final [BSA] (mg/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>49</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>48.5</td>
<td>1.5</td>
<td>0.3</td>
</tr>
<tr>
<td>48</td>
<td>2</td>
<td>0.4</td>
</tr>
<tr>
<td>47.5</td>
<td>2.5</td>
<td>0.5</td>
</tr>
<tr>
<td>47</td>
<td>3</td>
<td>0.6</td>
</tr>
<tr>
<td>46.5</td>
<td>3.5</td>
<td>0.7</td>
</tr>
<tr>
<td>46</td>
<td>4</td>
<td>0.8</td>
</tr>
</tbody>
</table>

2.3.2.2 BCA Assay

The protein samples were removed from -80°C and thawed on ice. The samples were prepared by mixing 36 µL of MQ-H₂O with 4 µL of protein sample and then aliquoting 20 µL into each of two wells. The rest of the assay was carried out as per the manufacturer’s instructions. The measured absorbances at 570 nm were plugged into the equation obtained from the standard curve in order to determine the total protein concentration of the samples.

2.3.3 Western Blot

2.3.3.1 Separating Gel

A 1 mm, 10% SDS-PAGE gel was used to separate the target proteins isolated from the transfected cells, which fall in the 20-70 kDa range. The gel consisted of 10% acrylamide, 0.33% Bis solution, 0.1% SDS (w/v), 0.375 M Tris-HCl pH 8.8, 0.1% ammonium persulfate (APS) (w/v), and 0.05% TEMED in dH₂O. The APS was made fresh and it, along with the TEMED, was added just prior to pouring the gel. Once poured, the gel was covered with distilled water to prevent it from drying out during polymerization.
2.3.3.2 Stacking Gel
Prior to separation, the isolated proteins were first stacked using a 5% gel. This was made up of 5% acrylamide, 0.17% Bis solution, 0.1% SDS (w/v), 0.125 M Tris pH 6.8, 0.1% APS, and 0.125% TEMED in dH$_2$O. The distilled water was poured off of the top of the separating gel and, as before, the APS and TEMED were added just prior to pouring the stacking the gel. Once layered on top of the separating gel, a comb was inserted into the stacking gel and it was left to polymerize.

2.3.3.3 Sample Preparation
Based on the results of the BCA assay, samples were prepared to contain a total of 40 µg of protein, to which 6 µL of 4x sample buffer were added to obtain a final volume of 24 µL. The samples were then centrifuged briefly, heated at 99°C for 10 minutes in a heat block, and then centrifuged again briefly before being stored at -80°C until use.

2.3.3.4 SDS-PAGE Gel
The samples to be electrophoresed were thawed and briefly centrifuged before loading. The gel was placed in a Mini-PROTEAN 3 tank (Bio-Rad Laboratories, USA) and submerged in 1 x SDS running buffer. The samples were loaded along with 1.2 µL of MagicMark XP Western Protein Standard (Invitrogen, USA) and 5 µL of PageRuler Prestained Protein Ladder (Fermentas, Canada) to act as size markers for the SDS-PAGE gel and western blot, respectively. The gel was run at a constant voltage of 120 V until the dye front reached the bottom of the gel.

2.3.3.5 Transfer
Proteins were transferred from the gel to a nitrocellulose membrane (Hybond-C Extra, GE Healthcare, UK) using the wet transfer method. The gel and the nitrocellulose membrane were sandwiched in a cassette with three layers of
filter paper and a foam pad on either side and placed in a Mini Trans-Blot cell system (Bio-Rad, USA). The transfer was run in ice-cold transfer buffer at 220 mA for 2.5 hours on ice. Once completed, the membrane was removed and rinsed in PBST and then the efficiency of the transfer was verified using Ponceau stain. Upon successful transfer, the membrane was blocked for one hour at room temperature in 2.5 % skim milk in PBST on a rotator.

### 2.3.3.6 Development

Following blocking, the membrane was incubated with the desired primary (1°) antibody overnight at 4°C on a rotator. The next day the blot was washed four times for 5 minutes in PBST and then incubated for 2 hours at room temperature with the appropriate secondary (2°) antibody. The blot was washed again four times for 5 minutes in PBST before being treated with a 1:1 mixture of freshly prepared SuperSignal West Pico Stable Peroxide Solution and Luminol/Enhancer Solution (Thermo Scientific, USA) for 5 minutes. The excess developing solution was removed and the blot was transferred to a cassette where it was sandwiched between two sheets of transparency film and exposed to Kodak Medical X-ray Film MXB. The sizes of the proteins were determined by comparing their positions relative to those of the molecular mass standards in the PageRuler ladder. After exposure, the membrane was washed for 1 hour in fresh PBST and then incubated with the next primary antibody at 4°C overnight on a rotator. The same process was repeated the following day in order to obtain an x-ray film showing the presence of the second target protein. Finally, the whole procedure was repeated once more with a primary antibody targeting a housekeeping protein to be used as a loading control in order to verify equal loading between all the samples. The antibodies used in this study are listed in Table 2-6.
Table 2-6: Antibodies used for western blotting

<table>
<thead>
<tr>
<th>1° Antibody (Manufacturer)</th>
<th>Concentration</th>
<th>2° Antibody*</th>
<th>Dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAX8 (Biopat, Italy)</td>
<td>1:20,000</td>
<td>anti-rabbit IgG</td>
<td>1:10,000</td>
</tr>
<tr>
<td>E2F1** (Santa Cruz Biotechnology, USA)</td>
<td>1:200</td>
<td>anti-rabbit IgG</td>
<td>1:2,500</td>
</tr>
<tr>
<td>β-actin** (Abcam, USA)</td>
<td>1:10,000</td>
<td>anti-mouse IgG</td>
<td>1:5,000</td>
</tr>
<tr>
<td>GAPDH (Santa Cruz Biotechnology, USA)</td>
<td>1:1,000</td>
<td>anti-goat IgG</td>
<td>1:5,000</td>
</tr>
</tbody>
</table>

* Secondary antibodies were labeled with horseradish peroxidase (HRP) and manufactured by Sigma-Aldrich, USA
** Provided by Prof. Antony Braithwaite (University of Otago, NZ)

2.3.4 Densitometry

In cases where the loading control was not equal between the samples, the density of each of the target protein bands was measured using the GS-700 Imaging Densitometer (Bio-Rad, USA) and the associated Quantity One software (version 4.6.3). The density of the background was subtracted and then the ratio of the target protein band to the loading control was calculated. These values were used to plot a graph in order to be able to compare the amount of protein present in the different samples.
2.4 Primers

All of the primers used in this study are listed in Table 2-7. The GL3F2 and GL3R1 primers were kindly gifted from Dr. Phil Daniel (University of Otago, NZ). The remaining primers were designed using Primer3 software and their properties were analyzed with the online IDT Oligo Analyzer tool. Custom primers were obtained from Invitrogen, USA.

**Table 2-7:** Experiment-specific primers used in this study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ =&gt; 3’)</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>GL3F2</td>
<td>CTGCATTCTAGTTGTGGTTTGTCC</td>
<td>Forward</td>
</tr>
<tr>
<td>GL3R1</td>
<td>GCAGTTGCTCTCCAGGGTTCC</td>
<td>Reverse</td>
</tr>
<tr>
<td><strong>Chromatin Immunoprecipitation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAX8 binding site 1-F</td>
<td>CGCGTTAAAAGCCAATAGGAA</td>
<td>Forward</td>
</tr>
<tr>
<td>PAX8 binding site 1-R</td>
<td>TTTACGCGCACAATCCTTT</td>
<td>Reverse</td>
</tr>
<tr>
<td>PAX8 binding site 1B-F</td>
<td>CGCGTTAAAAGCCAATAGGAA</td>
<td>Forward</td>
</tr>
<tr>
<td>PAX8 binding site 1B-R</td>
<td>AAAGTCCCGCCACTTTTAC</td>
<td>Reverse</td>
</tr>
<tr>
<td>PAX8 binding site 2-F</td>
<td>CGCACCTATAGAAAGGTCAGTGG</td>
<td>Forward</td>
</tr>
<tr>
<td>PAX8 binding site 2-R</td>
<td>AGGCTTTGTCCCCGATGGTA</td>
<td>Reverse</td>
</tr>
<tr>
<td>PAX8 binding site 2B-F</td>
<td>AGAAAGGTCAGGTCAGTGGCG</td>
<td>Forward</td>
</tr>
<tr>
<td>PAX8 binding site 2B-R</td>
<td>AGGCTTTGTCCCCGATGGCA</td>
<td>Reverse</td>
</tr>
<tr>
<td>ChIP site 2-F</td>
<td>TATGTTCCGGTGTCCCCAACGC</td>
<td>Forward</td>
</tr>
<tr>
<td>ChIP site 2-R</td>
<td>CAACGCGCGCGGGTTTCATT</td>
<td>Reverse</td>
</tr>
<tr>
<td>PAX8 binding site 3-F</td>
<td>GATGGAAGAAGGTGCTGATG</td>
<td>Forward</td>
</tr>
<tr>
<td>PAX8 binding site 3-R</td>
<td>AGTGCAGGGTCAAGACAGAGC</td>
<td>Reverse</td>
</tr>
<tr>
<td>PAX8 binding site 4-F</td>
<td>CTGGGTAGGATAGAAGCTGTATG</td>
<td>Forward</td>
</tr>
<tr>
<td>PAX8 binding site 4-R</td>
<td>TCAACCTGTAGCCCCCAAC</td>
<td>Reverse</td>
</tr>
<tr>
<td>PAX8 binding site 5-F</td>
<td>TCAAGCAGATGCTATACGGTGG</td>
<td>Forward</td>
</tr>
<tr>
<td>PAX8 binding site 5-R</td>
<td>TAATCCTGCAACCAGCCCTA</td>
<td>Reverse</td>
</tr>
<tr>
<td>PAX8 binding site 6-F</td>
<td>TAGGGGCCACAGAATGAGAAG</td>
<td>Forward</td>
</tr>
<tr>
<td>PAX8 binding site 6-R</td>
<td>CTGCAAGTCCATCCAGCCCTA</td>
<td>Reverse</td>
</tr>
<tr>
<td>E2F1 Promoter Control-F</td>
<td>CCTCTGGCTTAGCCTGTTT</td>
<td>Forward</td>
</tr>
<tr>
<td>E2F1 Promoter Control-R</td>
<td>CTCCAATGCTGACCACACAC</td>
<td>Reverse</td>
</tr>
</tbody>
</table>
Table 2-7 continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ =&gt; 3’)</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deletion Constructs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2F1 Promoter F1</td>
<td>TAATATGCTAGCTGGGACTTGAGAAAAG</td>
<td>Forward</td>
</tr>
<tr>
<td>E2F1 Promoter F2</td>
<td>TTATATGCTAGCCTGGTGAGAATG</td>
<td>Forward</td>
</tr>
<tr>
<td>E2F1 Promoter F3</td>
<td>TAATATGCTAGCTGAAGCTGAGCTTTG</td>
<td>Forward</td>
</tr>
<tr>
<td>E2F1 Promoter F4</td>
<td>TTATATGCTAGCTGAAGCTCTGAGCTTTG</td>
<td>Forward</td>
</tr>
<tr>
<td>E2F1 Promoter F5</td>
<td>TTATATGCTAGCTGAAGCTCTGAGCTTTG</td>
<td>Forward</td>
</tr>
<tr>
<td>E2F1 Promoter F6</td>
<td>TTATATGCTAGCTGAAGCTCTGAGCTTTG</td>
<td>Forward</td>
</tr>
<tr>
<td>E2F1 Promoter R</td>
<td>AATAATAAGCTTCAGTACCAGGAATCGCAATAGG</td>
<td>Reverse</td>
</tr>
<tr>
<td>E2F1 Promoter R2</td>
<td>TTATATAAGCTTCAGGAGATTCGAGACGAGA</td>
<td>Reverse</td>
</tr>
<tr>
<td><strong>Site-Directed Mutagenesis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 3 Mutagenic F1</td>
<td>GAGGTACCGAGCTCTTAAGGCAACGAGATGAG</td>
<td>Forward</td>
</tr>
<tr>
<td>Site 3 Mutagenic R1</td>
<td>AGCCTCTCTTCTTCTGACACTGACCAACGACGTAGG</td>
<td>Reverse</td>
</tr>
<tr>
<td>Site 3 Mutagenic F2</td>
<td>CAGGTTGAGGGGAAGTTGCAGAAGAGAGCTCTG</td>
<td>Forward</td>
</tr>
<tr>
<td>Site 3 Mutagenic R2</td>
<td>AACAGTACCGGAATGCAAAGGCGAGGATGAG</td>
<td>Reverse</td>
</tr>
</tbody>
</table>

2.5 Chromatin Immunoprecipitation

2.5.1 Binding Site Prediction

Sites predicted to bind PAX8 in the 728 bp region upstream of the transcription start site of the E2F1 gene promoter were identified using the online binding site prediction tool ConTra (Hooghe et al., 2008). The alignments selected were the Ensembl PECAN 12 amniota vertebrates and the UCSC multiz 17way 5000 upstream, with a stringency set to core = 0.90 and similarity matrix = 0.75. The position weight matrix (PWM) chosen was V$PAX8_01$, from the TRANSFAC database, which has the sequence NNNTNNNGNTGAN (Matys et al., 2003).
2.5.2 Cells

A498 cells were grown to a confluence of 80-95% in a T-175 flask and then harvested by trypsinization as described in Section 2.2.1. Following centrifugation, the cell pellet was gently resuspended in 9 mL of media. The cells were seeded into eight 10 cm$^2$ cell culture dishes, each containing 10 mL of media to which 1 mL of cell suspension was added drop-wise. The plates were tapped on four sides to ensure an even distribution of the cells. The cells were then cultured until they reached a confluence of at least 80%, which usually took approximately 48 hours.

2.5.3 Crosslinking

Formaldehyde (~38% w/w) was added drop-wise to a final concentration of 1% to each plate and the plates were swirled to evenly distribute the solution. This was followed by a 10 minute incubation at room temperature allowing the formaldehyde to crosslink the proteins to the DNA. Glycine (2.5 M) was then added drop-wise to each plate to a final concentration of 0.125 M followed by swirling to evenly distribute the solution and a 5 minute incubation at room temperature to quench the crosslinking reaction. The mixture was then discarded and the cells were washed two times in cold, sterile PBS and put on ice.

2.5.4 Sonication

The cells were scraped from all of the plates into 1.5 mL of SDS lysis buffer and then divided into 350 μL aliquots and kept on ice. Each sample was sonicated ten times for 15 seconds at 25% amplitude using a Vibra-Cell VCX130 sonicator (Sonics and Materials, Inc., USA). The success of the sonication was verified prior to continuing by pooling 1 μL from each aliquot and mixing the pooled sample with an equal volume of TE buffer. The crosslinks were reversed with the addition of 1μL of RNase A (10 mg/mL) and 0.5 μL of proteinase K (10 mg/mL) followed by incubation at 37°C for 1.5 hours in a shaking heating block at 1,000 rpm. A 5 μL aliquot of the sample
was then mixed with 2 µL of xylene-cyanol loading dye and electrophoresed at 100 V on a 1.5% agarose gel until the dye front migrated three-quarters of the way down the gel to confirm that the chromatin fragments were in the target size range of 200 – 1,000 bp.

2.5.5 Pre-clearing

The sonicated chromatin aliquots were centrifuged for 10 minutes at 12,000 x g and 4°C and the supernatants pooled together in a 15 mL Falcon tube. The pooled sample was diluted 2.5 times with ChIP dilution buffer. A 70 µL sample of Dynabeads Protein G (Invitrogen Dynal AS, Norway) was obtained, the buffer removed, and the beads resuspended in diluted chromatin before being added to the sample. The diluted chromatin was then incubated with the beads for 1 hour at 4°C on a rotator at 4.5 rpm.

2.5.6 Immunoprecipitation

The beads were removed from the pre-cleared chromatin, which was then divided into two aliquots. One of the aliquots was incubated with 4 µg of anti-PAX8 antibody, while the other was left untreated in order to act as the no antibody control. Both samples were incubated at 4°C on a rotator at 5 rpm overnight. The beads were washed twice with PBS and then stored in 70 µL of PBS in the fridge for use the next day. Following overnight incubation, a 50 µL aliquot of fresh beads was added to the PAX8 sample and 50 µL of the 70 µL of beads used for pre-clearing was added to the no antibody control. The samples were then incubated with the beads at 4°C on a rotator at 5 rpm for 4 hours. The beads were collected from each of the samples and then washed two times with each of four wash buffers: low salt, high salt, LiCl, and TE. Each of the wash steps involved resuspending the beads in 1 mL of the desired wash buffer and putting the beads on a rotator at 4°C for 10 minutes. The supernatant containing unbound chromatin from the no antibody control was saved for use as the input DNA sample (Section 2.5.8).
2.5.7 ChIP-DNA Elution

After the final TE wash, 50 µL of 10% Chelex-100 was added to the beads, vortexed at maximum speed for 10 seconds and incubated in a shaking heat block at 99°C and 1,000 rpm for 10 minutes. The tubes were cooled to room temperature and 2 µL of proteinase K (10 mg/mL) was added. The samples were vortexed at maximum speed for 10 seconds and incubated at 55°C and 1,300 rpm for 30 minutes. Following another 10 seconds of vortexing at maximum speed, the samples were incubated at 99°C and 1,000 rpm for a further 10 minutes in order to inactivate the proteinase K. The samples were then centrifuged at 16,100 x g for 2 minutes and the supernatant containing the eluted DNA was transferred to a fresh microfuge tube. A 50 µL amount of sterile water was added to the beads/Chelex-100 mixture, vortexed for 10 seconds, and centrifuged for 2 minutes at 16,100 x g. The supernatant was again removed and pooled with the previous aliquot. The eluted DNA was quantified by measuring the absorbance at 260 nm using a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, USA) and stored at -80°C until needed.

2.5.8 Input DNA Elution

The supernatant removed from the no antibody control beads was used as the input DNA sample (Section 2.5.6). The sample was mixed with three times the volume of 100% ethanol and incubated at -80°C for 30 minutes. The solution was then aliquoted into 1.5 mL tubes and centrifuged at 16,100 x g and 4°C for 10 minutes. The pellets were collected by using 1 mL of 70% ethanol to flush the sides of each tube and then pooled together in a single microfuge tube. The solution was centrifuged at 16,100 x g for 5 minutes at 4°C and then the pellet was dried. A 400 µL aliquot of 10% Chelex-100 was added to the dried pellet, vortexed for 10 seconds, and then placed on a rotator at 20 rpm and 4°C until the pellet was completely resuspended. Once resuspended, the sample was processed as outlined in Section 2.5.7.
2.5.9 **Quantitative PCR**

The quantitative PCR (qPCR) reactions were carried out using the ABI 7300 Real-Time PCR System (Applied Biosystems, USA). The reaction mixture contained the desired amount of DNA, primers, Platinum SYBR Green qPCR SuperMix-UDG with ROX (Invitrogen, USA) and was made up to a final volume of 20 µL with sterile MQ-H2O. Each reaction was performed in duplicate using the following settings: 50°C for 2 minutes, 95°C for 10 minutes, 40 cycles of 95°C for 15 seconds and 60°C for 1 minute, and an additional dissociation step of 95°C for 15 seconds, 60°C for 1 minute and 95°C for 15 seconds.

2.5.9.1 **Amplification Efficiency**

The amplification efficiency of each primer set was determined by generating a standard curve of the Ct value versus the log of the concentration of a set of DNA standards. The qPCR reactions were performed with DNA concentrations of 2, 3, 5, 10, 25, and 50 ng, each in a 1 µL aliquot. The rest of the reaction mixture consisted of 10 µL of the Platinum SYBR Green qPCR mix, 200 nM or 400 nM final concentration of forward and reverse primers (depending on the primer set), and was made up to 20 µL with sterile MQ-H2O. The amplification efficiency was used to normalize the ChIP-DNA signal to the input DNA signal.

2.5.9.2 **ChIP qPCR**

The input, PAX8 and no antibody control samples were all prepared to a concentration of 12.5 ng/µL with 4 µL being used for each qPCR reaction (50 ng total DNA). As with the amplification efficiency qPCR, the reaction mixture also contained 10 µL of Platinum SYBR Green qPCR mix, forward and reverse primers to a final concentration of 200 or 400 mM each, and was made up to a final volume of 20 µL with sterile MQ-H2O.
2.5.9.3 Amplicon Sequencing

The ChIP-qPCR products were run on a 1.5% agarose gel to confirm the presence of a single band. The target bands were excised and the DNA was extracted from the gel using the Qiagen Gel Extraction Kit (Qiagen, Germany). Cycle sequencing was performed using the following reaction components: 5 µL of gel purified DNA, 1 µL of BigDye Terminator (Applied Biosystems, USA), 3.5 µL of 5x sequencing dilution buffer (Applied Biosystems, USA), 1 µM of sequencing primer, and MQ-H$_2$O to a final volume of 20 µL. The reaction was carried out under the following conditions: 96°C for 1 minute 15 seconds and 25 cycles of 96°C for 45 seconds, 50°C for 45 seconds, and 60°C for 3 minutes 30 seconds.

Post-sequence clean-up was carried out by adding 2 µL of 3M sodium acetate, 1 µL of 125 mM EDTA, and 50 µL of room temperature 100% ethanol. The samples were vortexed and left at room temperature for 15 minutes in order to precipitate the extension products. The samples were then centrifuged at 16,100 x g for 20-30 minutes. The supernatant was carefully removed and discarded and the pellet was rinsed with 200 µL of 70% ethanol. The pellet was vortexed and centrifuged at 16,100 x g for 5 minutes. The supernatant was again discarded and the wash step repeated. Finally, the pellets were air dried and sent to the Anatomy Department for dideoxy sequencing using the 3730xl DNA Analyzer (Applied Biosystems, USA).

2.5.10 Data Analysis

The qPCR data was analyzed using a modified version of the ChIP-qPCR Primer Assay Data Analysis Template (SABiosciences, Germany). The assay site IP fold enrichment was determined for each predicted binding site using the equation: $2^{\Delta\Delta Ct}$, where $\Delta Ct = (Ct(IP) – Ct(input x dilution factor)) – (Ct(no antibody control) – Ct(input x dilution factor))$. This value was then calibrated to the value of the control site in order to get the fold change in site.
occupancy. The entire ChIP experiment was repeated four times and the standard error of the mean was calculated.

2.6 Deletion Constructs

2.6.1 PCR

The target sequences for the deletion constructs were amplified from the pE2F1(-728) promoter construct. The reaction mixture contained 10 ng of plasmid DNA, 1 μM each of forward and reverse primers, 5% DMSO, 1.0 U Phusion Hot Start II DNA Polymerase (Finnzymes, Finland), 1x buffer HF (Finnzymes, Finland), and MQ-H2O to a final volume of 50 μL. The PCR was then carried out using the following settings: 98°C for 2 minutes, 30 cycles of 98°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute, and a final extension step of 72°C for 7 minutes.

2.6.2 Agarose Gel Electrophoresis

5 μL of PCR product was mixed with 2 μL of 6x bromophenol blue DNA loading dye. A 1.5% agarose gel was submerged in 1x TAE buffer containing 25 μg/L of ethidium bromide in a Mini-Sub Cell GT Cell (Bio-Rad, USA). Electrophoresis was carried out at 100 V until the dye front had run at least three-quarters of the way down the gel. The sizes of the bands were verified by comparison to those contained in the 1 Kb Plus DNA ladder (Invitrogen, USA) that was run alongside the samples.

2.6.3 DNA Extraction

The target insert DNA bands were extracted from the agarose gel using the Qiagen Gel Extraction kit (Qiagen, Germany) according to the manufacturer’s instructions. Briefly, the band of interest was excised from the agarose gel using a razor blade and placed in a microfuge tube. The gel was weighed and three times the volume of Buffer QG was added. The sample was incubated at 50°C.
for 10 minutes with occasional vortexing to completely dissolve the gel. An equal volume of isopropanol was added and mixed and then the sample was loaded onto a column and centrifuged for 1 minute to bind the DNA to the column. The bound DNA was washed with Buffer PE and centrifuged again for 1 minute. Finally, the DNA was eluted from the column using Buffer EB and centrifuging for 1 minute into a fresh tube.

### 2.6.4 Preparation of Insert and Vector DNA

The insert and vector DNA were prepared for ligation by doing a double digest with restriction enzymes HindIII and NheI. The reaction was carried out for 16 hours (overnight) at 37°C and then the enzymes were inactivated by heating the mixtures at 65°C for 20 minutes. The reaction components included 30 μL of DNA, 0.5 μL of HindIII, 0.5 μL of NheI, 0.5 μL of BSA, 5 μL of NEB2 buffer and 13.5 μL of MQ-H₂O.

### 2.6.5 Ligation

Ligations were carried out at 14°C for 24 hours using an insert to vector ratio of 3:1. A concentration of 30 fmol of vector ends and 90 fmol of insert ends was calculated using the following equations:

\[
\begin{align*}
\text{Vector DNA} & = 30 \text{ fmol x 1 } \mu \text{g/3000 fmol x size of vector (bp)/1000 bp} \\
\text{Insert DNA} & = 90 \text{ fmol x 1 } \mu \text{g/3000 fmol x size of insert (bp)/1000 bp}
\end{align*}
\]

The insert and vector DNA were mixed with 4 μL of 5x buffer, 0.1 unit of T4 DNA ligase (both Gibco Brl Life Technologies, USA), and made up to 20 μL with MQ-H₂O. The T4 DNA ligase was then inactivated by heating the reaction mixtures at 70°C for 10 minutes.
2.6.6 Transformations

The newly ligated plasmid DNA was transformed into competent DH5α cells. 5 μL of the ligation mix was added to a 100 μL aliquot of DH5α cells, mixed gently, and incubated for 30 minutes on ice. The cells were then heat-shocked for 90 seconds at 42°C followed by a 1-2 minute incubation on ice. The transformed cells were plated on LB plates containing ampicillin and incubated at 37°C overnight. Untransformed DH5α cells and cells transformed with digested vector were used as controls.

2.6.7 Single Colony PCR

Several single colonies were selected and individually inoculated into 500 μL of LB + Amp and grown at 37°C in a 170 rpm shaker for a minimum of 4 hours. A 5 μL aliquot of each culture was then transferred to a PCR tube, heated for 2 minutes at 94°C and then cooled on ice for a minimum of 5 minutes prior to the addition of the PCR master mix. The reaction was made up to 20 μL and cycled through the following reaction conditions: 95°C for 2 minutes, 30 cycles of 95°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute, then 72°C for 7 minutes. The PCR products were run on 1.5% agarose gel in order to confirm the presence of the insert.

2.6.8 Minipreparation of Plasmid DNA

The remainder of the 500 μL culture of colonies containing the desired insert were inoculated into 100 mL of LB+Amp and grown overnight at 37°C and 170 rpm. 1 mL of the culture was centrifuged at 16,100 x g for 1 minute. The supernatant was discarded and the pellet was resuspended in 100 μL of lysis buffer (plus RNase). 200 μL of freshly made alkaline SDS was then added and the sample was left on ice for 5 minutes. Next, 150 μL of 3M sodium acetate, pH 5 was added, the sample vortexed, and left on ice for 30 minutes. The samples were then centrifuged at 16,100 x g for 10 minutes in a microfuge. The supernatant was transferred to a fresh tube to which 1 mL of 100% ethanol was added, vortexed, and placed at -70°C for 15-20 minutes. The samples were then
centrifuged at 16,100 x g for 10 minutes. The supernatant was discarded and the pellet washed with 500 µL of 95% ethanol. The pellets were centrifuged again for 5 minutes at 16,100 x g and 4°C. The supernatant was discarded and the pellet was dried prior to being resuspended in sterile MQ-H$_2$O.

2.6.9 Sequencing

The deletion construct inserts were sequenced as outlined in Section 2.5.9.3.

2.6.10 Midipreparation of Plasmid DNA

Plasmid DNA with the correct sequence was purified via midiprep using the PureLink HiPure Plasmid Filter Midiprep Kit (Invitrogen, USA) as per the manufacturer’s instructions. Briefly, an overnight culture of transformed DH5α cells was centrifuged at 4,000 x g for 10 minutes at 4°C. The media was removed and the cell pellet was resuspended in resuspension buffer containing RNase A. Lysis buffer was then added, mixed until homogeneous, and incubated at room temperature for 5 minutes. Precipitation buffer was added and mixed thoroughly and then the mixture was centrifuged at 12,000 x g for 10 minutes at room temperature. The supernatant was loaded onto an equilibrated manufacturer-supplied anion-exchange column, drained, and then the column was washed twice with wash buffer. The DNA was eluted into a fresh tube with elution buffer. Isopropanol was added to the eluted DNA, mixed, and centrifuged at 12,000 x g for 30 minutes at 4°C. The supernatant was discarded and the pellet was air-dried before being resuspended in 100 µL of sterile MQ-H$_2$O and stored at -20°C.

2.6.11 Promoter-Reporter Assay

The resulting deletion constructs were used to transfect cells (Section 2.2.2.1) for analysis in the promoter-reporter assay, as described in Section 2.2.3.
2.7 Site-Directed Mutagenesis

2.7.1 Primer Design

Two sets of primers were designed to incorporate three mutations into the predicted site 3 PAX8 binding site at locations of high abundance bases. The internal primers, R1 and F2, contained the desired substitutions, while the external primers, F1 and R2, included the recognition sequences for the restriction enzymes SacI and HindIII, respectively.

2.7.2 PCR

Three separate reactions were carried out in order to obtain the full-length insert with the desired mutations, as shown in Figure 2-1. The first reaction amplified the target region of the pE2F1(-728) construct from the 5’ end to the predicted binding site using primers F1 and R1 with the following reaction components: 0.2 mM dNTPs, 0.5 μM of each primer, approximately 10 ng of DNA, 0.02 U/μL of Phusion High-Fidelity DNA Polymerase and 1x HF Buffer (both Finnzymes, Finland), and made up to 20 μL with MQ-H2O. The second reaction amplified the region from the predicted binding site to the 3’ end with primers F2 and R2 and the same reaction components as reaction 1 with the addition of 3% DMSO. The third reaction was performed in order to join the two overlapping fragments obtained from the first two reactions into the full-length mutant insert. The products from reactions 1 and 2 were each diluted 100-fold and then 1 μL of each was used as the template DNA for a reaction with primers F1 and R2, along with the same components as for reaction 2. Each of the three reactions were carried out under the following conditions: 98°C for 30 seconds, 30 cycles of 98°C for 10 seconds followed by 72°C for 15 seconds, and a final 7 minutes at 72°C.
Figure 2-1: Site-directed mutagenesis reactions
The three reactions required to obtain the full-length piece of DNA with the desired mutations incorporated are shown. The products of reactions 1 and 2 were used as the template DNA for reaction 3 in order to obtain the mutations on both strands of the DNA sequence.

2.7.3 Cloning
The mutated insert DNA was then cloned into the pGL2-basic vector as described in Sections 2.6.2 – 2.6.8 and sequenced as outlined in Section 2.5.9.3.

2.7.4 Promoter-Reporter Assay
The successfully cloned mutant construct was used to transfect cells (Section 2.2.2.1) for analysis in the promoter-reporter assay as described in Section 2.2.3.
CHAPTER 3
– PAX8 Activation of the $E2F1$ Promoter –

3.1 Introduction

PAX8 is expressed in a variety of different cancer types (Kroll et al., 2000; Pellizzari et al., 2006; Bowen et al., 2007; Tong et al., 2008; Tong et al., 2009) although its function remains undetermined. As described in Section 1.4, previous research using microarray analysis to examine gene expression following PAX8 knockdown in A498 renal cell carcinoma cells identified a resulting downregulation of several genes involved in the cell cycle, including $E2F1$ (Li et al., 2011). Subsequent western blot analysis showed a downregulation of the E2F1 protein following the treatment of several cancer cell lines (including A498) with siRNA targeting PAX8 (Li et al., 2011). $E2F1$ is known as an important regulator of the cell cycle and is often overexpressed in cancer. This, combined with the results of the PAX8 knockdown experiments, suggests that PAX8 may be involved in the regulation of the $E2F1$ gene promoter.

The purpose of this chapter is to further examine the relationship between PAX8 and $E2F1$ by characterizing the effects of PAX8 knockdown and overexpression on the activity level of the $E2F1$ promoter.

3.2 $E2F1$ Promoter Constructs

Two plasmids containing 242 bp and 728 bp of the $E2F1$ promoter sequence upstream of the transcription start site were obtained from Assoc. Prof. Masa-Aki Ikeda and Prof. Kiyoshi Ohtani, respectively. The constructs were generated by Johnson and colleagues (1994) who cloned the $E2F1$ promoter segments into the pGL2-basic vector, which contains the luciferase reporter gene, thereby
allowing the activity of the promoter segments to be measured. The plasmid maps of both constructs, identifying the locations that the inserts were cloned into the vector, are shown in Figure 3-1.

**Figure 3-1: E2F1 Promoter Constructs**

Each of the E2F1 promoter segments was cloned directly upstream of the luciferase gene (luc) in the multiple cloning site of the pGL2-basic vector. The 242 bp insert was cloned using SmaI and HindIII, while the 728 bp insert was cloned using SacI and HindIII. The HindIII end of the inserts was first digested with Styl and then filled in with HindIII. The arrow at position +1 indicates the transcription start site of the promoter and each segment contained an additional 70 bp downstream of this site.
3.2.1 Sequence Validation

Prior to using either of the E2F1 promoter constructs in any experiments, the first step was to validate that they did contain the correct E2F1 promoter sequences. This was accomplished by first amplifying the inserts by PCR using two primers (GL3F2 and GL3R1) that recognize sequences in the pGL2-basic vector, just outside of the multiple cloning site. The amplified regions were then sequenced and aligned with the E2F1 reference sequence using ClustalW2 (EMBL-EBI, UK). The alignments for the two promoter constructs with the reference sequence are shown in Figure 3-2. There is a section in the middle of the 728 insert that failed to be fully sequenced, despite several attempts. Given that this particular section has a very high GC content, it is possible that this resulted in some issues during the sequencing reaction. It was decided to proceed under the assumption that the pE2F1(-728) construct did indeed contain the entire segment of the E2F1 promoter seeing as the rest of the insert was successfully sequenced. The 242 insert was also successfully sequenced thus validating the pE2F1(-242) construct.
Figure 3-2: E2F1 promoter construct sequence alignment

Multiple sequence alignment of the E2F1 promoter segments with the reference sequence in ClustalW2 (EMBL-EBI, UK). Asterisks indicate exact matches and dashes indicate gaps in the sequence.
3.3 PAX8 Knockdown

To examine the effects of PAX8 knockdown on the activity of the E2F1 promoter, the set-up of the microarray experiments was replicated whereby PAX8 was knocked down in the presence of one of the E2F1 promoter constructs. The K1 thyroid cancer cell line was selected for this experiment as the cells had previously been determined to be able to withstand the cytotoxic effects of the siRNA/DNA co-transfection (Li, 2010).

The K1 cells were first treated with siRNA targeting PAX8 (siPAX8) or a non-targeting control (siControl) and were co-transfected 24 hours later with one of the E2F1 promoter constructs as well as the pCMV-β-gal expression vector and the pCR3.1 plasmid for a total of 250 ng of DNA (Table 3-1).

<table>
<thead>
<tr>
<th></th>
<th>pGL2-basic</th>
<th>pE2F1(-242)</th>
<th>pE2F1(-728)</th>
</tr>
</thead>
<tbody>
<tr>
<td>siControl</td>
<td>100 ng</td>
<td>100 ng</td>
<td>100 ng</td>
</tr>
<tr>
<td>siPAX8</td>
<td>100 ng</td>
<td>100 ng</td>
<td>100 ng</td>
</tr>
<tr>
<td>pCR3.1</td>
<td>100 ng</td>
<td>100 ng</td>
<td>100 ng</td>
</tr>
<tr>
<td>pCMV-β-gal</td>
<td>50 ng</td>
<td>50 ng</td>
<td>50 ng</td>
</tr>
</tbody>
</table>

The cells were harvested 48 hours post-DNA transfection and the activity of the promoter constructs was measured using the luciferase and β-gal assays. As seen in Figure 3-3, there was a decrease in the activity of both E2F1 promoter constructs following PAX8 knockdown as compared with the control siRNA. The fold difference for the pE2F1(-242) construct was 1.4 (P > 0.05) while that of the pE2F1(-728) construct was 3.2 (P < 0.05). This suggests that PAX8 is
involved in regulating the *E2F1* promoter and that there is a greater effect in the longer pE2F1(-728) construct than in the shorter pE2F1(-242).

**Figure 3-3: E2F1 promoter activity following PAX8 knockdown**

The relative luciferase activity of the two *E2F1* promoter constructs and the pGL2-basic vector (used as a control) following knockdown of PAX8 (siPAX8) or treatment with a non-targeting control siRNA (siControl). The values shown represent the mean +/- 1 SEM. *P < 0.05* (unpaired, two-tailed t test).

### 3.3.1 Validation of siRNA

To confirm that the siRNA was effectively targeting PAX8, K1 cells were harvested 72 hours post-siRNA transfection (to coincide with the time that they would be harvested for assay analysis) and the total protein was isolated. Analysis of the protein samples by SDS-PAGE gel and western blot showed that following treatment with siPAX8, there was a decrease in the amount of PAX8 protein present relative to that in the siControl-treated sample (Figure 3-4).
A decrease in the mRNA levels of *PAX8* following siPAX8 treatment had previously been observed in three RCC cell lines (A498, 786-O, and 769-P) as determined by RT-qPCR (Li, 2010). This, combined with the protein results, indicates that siPAX8 is successfully targeting and knocking down *PAX8*.

### 3.4 PAX8 Overexpression

Since the knockdown of PAX8 appears to decrease the activity of the *E2F1* promoter, it raises the question as to whether or not the inverse is also true and that overexpression of PAX8 would increase *E2F1* promoter activation. If this is the case, this could be one of the mechanisms through which PAX8 plays a role in cancer cells – by activating *E2F1* transcription and thus driving the cell cycle.

For these experiments, HEK-293 cells were selected as they do not express endogenous PAX8. The cells were transfected with a combination of pCMV5-Pax8 expression vector, one of the *E2F1* promoter constructs, pCMV-β-gal expression vector, and pCR3.1 (Table 3-2). The pCR3.1 plasmid served to bring the total amount of DNA up to 600 ng to keep the amount transfected constant between all of the samples. The activity levels of the promoter constructs were analyzed 48 hours post-transfection.
Table 3-2: Quantities of DNA transfected into HEK-293 cells to examine the effects of PAX8 overexpression on the E2F1 promoter

<table>
<thead>
<tr>
<th></th>
<th>pGL2-basic</th>
<th>pE2F1(-242)</th>
<th>pE2F1(-728)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCMV-Pax8 (ng)</td>
<td>0 100 200 400</td>
<td>0 100 200 400</td>
<td>0 100 200 400</td>
</tr>
<tr>
<td>Promoter Construct (ng)</td>
<td>75 75 75 75</td>
<td>75 75 75 75</td>
<td>75 75 75 75</td>
</tr>
<tr>
<td>pCR3.1 (ng)</td>
<td>500 400 300 100</td>
<td>500 400 300 100</td>
<td>500 400 300 100</td>
</tr>
</tbody>
</table>

Figure 3-5 indicates that increasing the amount of PAX8 transfected into the cells corresponds with an increase in the activity of the E2F1 promoter constructs. Despite the higher RLU values obtained by the pE2F1(-728) construct following transfection of 400 ng of PAX8 (782,777.628 +/- 382,705.095) as compared to those obtained by the pE2F1(-242) construct (221,058.820 +/- 41,205.497), there was a much smaller difference in the fold change relative to the 0 ng PAX8 sample between the two. The longer pE2F1(-728) construct showed a fold increase of 4.2 (P > 0.05) following transfection with 400 ng of PAX8 as compared to the shorter pE2F1(-242), with a fold change of 3.7 (P < 0.05).
The relative luciferase units of the two E2F1 promoter constructs and the pGL2-basic vector (used as a control) following co-transfection with either 0 ng, 100 ng, 200 ng or 400 ng of pCMV5-Pax8 expression vector. The values shown represent the mean +/- 1 SEM. * P < 0.05 (one-way ANOVA, Tukey’s test).

### 3.4.1 Verification of the PAX8 expression plasmid

To ensure that PAX8 was indeed being overexpressed in each of the treatments, cells were harvested 48 hours post-transfection and the total protein isolated. Analysis of the protein samples was done using SDS-PAGE and western blot, the results of which can be seen in Figure 3-6. Given that the loading control was unequal between the four samples, the protein spots were quantified using a densitometer.

Western blot analysis of PAX8 expression in HEK-293 cells following transfection with various amounts of pCMV5-Pax8 expression vector. GAPDH was used as a loading control.
The blot was scanned using the GS-700 Imaging Densitometer (Bio-Rad, USA) and analyzed with the associated QuantityOne software (v.4.63). The density of each protein band was measured using a box of the same size (Figure 3-7). An additional box was used to measure an unstained area to be used as the background value to be subtracted from each of the protein band densities. The adjusted density of each PAX8 protein band was then divided by the adjusted density of the corresponding GAPDH band in order to determine the relative protein concentration in each treatment sample (Figure 3-8).

**Figure 3-7**: Densitometer analysis of PAX8 expression

U1 = 400 ng PAX8, U2 = blank/background, U3 = 0 ng PAX8, U4 = 100 ng PAX8, and U5 = 200 ng PAX8.
Figure 3-8: PAX8 density ratios
The density ratio of PAX8 to GAPDH expression in each of the four different treatment samples as determined using densitometer analysis.

The results depicted in Figure 3-8 suggest that the expression of PAX8 increases from the first to the third treatment but then drops off in the final one. This is in partial agreement with the results of the western blot shown in Figure 3-6, where there is a visible increase from one treatment to the next. One possible explanation for this variation is that there may have been some error in the densitometry analysis of the 400 ng PAX8 sample. This is because the anti-goat secondary antibody used to detect the GAPDH also detected numerous other proteins resulting in a very strongly stained blot, making it difficult to isolate the PAX8 and GAPDH bands. Despite this, PAX8 is indeed being overexpressed in all three treatments as compared to the cells that were not transfected with the PAX8 expression vector (0 ng).

3.5 Discussion
The fact that the activity levels of both of the $E2F1$ promoter constructs were decreased following PAX8 knockdown and increased with PAX8 overexpression suggests that PAX8 is involved in the regulation of the $E2F1$ promoter. Whether this is accomplished via a direct interaction between the
protein and the gene or is the result of an indirect mechanism is yet to be
determined. The results of these two experiments also show that the activity of
the longer pE2F1(-728) construct was more greatly affected by both the
knockdown and overexpression of PAX8, with fold changes of 3.2 (P < 0.05)
and 4.2 (P > 0.05), respectively, as compared to the shorter pE2F1(-242)
construct, with fold changes of 1.4 (P > 0.05) and 3.7 (P < 0.05). This differing
level of responsiveness to PAX8 suggests that there may be one or more
additional factors involved in the regulation of the longer construct than the
shorter one. For example, if PAX8 is activating the $E2F1$ promoter by binding
to it directly, it is possible that more than one PAX8 binding site exists in the
promoter region and that an additional site or sites are located in the region
between -242 and -728 bp. The ability to bind a greater number of PAX8
proteins may then have an additive effect on the activation of the promoter.
Another possibility is that PAX8 interacts with a co-activator, whose binding
site may be located in the region between -242 and -728 bp. In this case, the
pE2F1(-242) construct would be missing the binding site of this co-activator
and would therefore not be able to be activated to the same level as the pE2F1(-
728) construct. If PAX8 is activating the $E2F1$ promoter via another
mechanism, such as by releasing autocrine growth factors that regulate $E2F1$
transcription, there may be an increased number of binding sites in the pE2F1(-
728) construct as compared to the shorter construct, thereby allowing for greater
transcriptional activation.
CHAPTER 4
– PAX8 Binds Directly to the E2F1 Promoter –

4.1 Introduction

In the previous chapter, the results of the promoter-reporter assay suggested that PAX8 was regulating the E2F1 promoter. The next step was, therefore, to determine whether this was being carried out via a direct interaction between the protein and the gene or if it was the result of an indirect mechanism. In order to accomplish this, chromatin immunoprecipitation (ChIP) was used to identify if and where PAX8 binds to the E2F1 promoter.

4.2 Binding Site Prediction

The first step in the process was to identify any sites predicted to bind PAX8 in the 728 bp region upstream of the E2F1 transcription start site, which is the portion contained in the long E2F1 promoter construct, pE2F1(-728). Seven potential PAX8 binding sites were identified, the last two being separated by only a single base. When the stringency was increased to core = 0.95 and similarity matrix = 0.85, three of the originally predicted sites were identified (site 2, site 4, and site 5), and sites 6 and 7 were reduced to a single site. The alignments of sites 1 through 5, and the sixth site that was identified at the higher stringency, are shown in Figure 4-1. The relative positions of the predicted sites in each of the two promoter constructs are shown in Figure 4-2.
Six potential PAX8 binding sites were identified in the *E2F1* promoter using the two alignment tools Ensembl PECAN 12 amniota vertebrates (12 species) and UCSC multiz 17way (13 species). The conserved sequences, with the human sequence acting as the template, are highlighted and labeled. Sites 1, 4, 5, and 6 are shown from the Ensembl PECAN 12 amniota vertebrates alignment results while sites 2 and 3 are shown from the UCSC multiz 17way alignment results.
Figure 4-1 continued
Figure 4-1 continued

The E2F1 promoter constructs contained 242 [pE2F1(-242)] or 728 bp [pE2F1(-728)] of the E2F1 promoter sequence upstream of the transcription start site, indicated by the arrow at +1. They also extended past the transcription start site for an additional 70 bp. The purple boxes indicate the relative positions of the six predicted PAX8 binding sites located in this particular region of the E2F1 promoter. Both promoter inserts were cloned into the pGL2-basic vector, which contains the firefly luciferase reporter gene (LUC).

Figure 4-2: Schematic representation of the E2F1 promoter constructs used in this study in relation to the E2F1 promoter
4.3 Quantitative PCR

4.3.1 Primer Design

Seven sets of primers were designed, one for each potential PAX8 binding site plus one for a region not predicted to bind PAX8, to act as a control. The sequences of the target sites and the amplicon lengths are listed in Table 4-1 with the positions of the amplicons relative to the predicted binding sites shown in Figure 4-3.

Table 4-1: ChIP target sequences and amplicon lengths

<table>
<thead>
<tr>
<th>Target</th>
<th>Sequence</th>
<th>Amplicon Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>CGTCACGGCCGCGGCCGCGG</td>
<td>131</td>
</tr>
<tr>
<td>Site 2</td>
<td>AGTCACACCACCGCGCCT</td>
<td>197</td>
</tr>
<tr>
<td>Site 3</td>
<td>GGTCACGTGCAGAAG</td>
<td>147</td>
</tr>
<tr>
<td>Site 4</td>
<td>ATGTCATGGGTGAGG</td>
<td>150</td>
</tr>
<tr>
<td>Site 5</td>
<td>GGGCTCTGGGTGAGCG</td>
<td>143</td>
</tr>
<tr>
<td>Site 6</td>
<td>TGAATGAGTGAGG</td>
<td>132</td>
</tr>
<tr>
<td>Control Site</td>
<td>N/A</td>
<td>144</td>
</tr>
</tbody>
</table>
Figure 4-3: Schematic representation of the locations of the regions to be amplified by the ChIP primers

The six predicted PAX8 binding sites in the 728 bp region, upstream of the transcription start site (position +1) of the E2F1 promoter are shown as purple boxes. The light blue boxes depict the areas to be amplified by each of the ChIP primer sets. The control site primers are located further upstream in a region that is not predicted to bind PAX8.

4.3.2 Reaction Conditions

The ideal primer concentration for each primer pair was determined. Each reaction was made up of 50 ng of DNA, Platinum SYBR Green qPCR SuperMix-UDG with ROX (Invitrogen, USA), MQ water, and forward and reverse primers. Primers for sites 4, 5, 6 and control were added at a final concentration of 400 nM, while site 3 primers were added to a final concentration of 200 nM (Figure 4-4).
The PCR products from reactions performed with the primer sets for sites 3-6 and the control both with 50 ng of template DNA and a non-template control (NTC). The spread of the bands is indicated by the bracket to the right of the gel. The sizes of the bands are as follows: Site 3 – 147 bp, Site 4 – 150 bp, Site 5 – 143 bp, Site 6 – 132 bp, Control Site – 144 bp. Lane 1, 1 Kb Plus DNA ladder. Lane 2, site 3. Lane 3, site 3 NTC. Lane 4, site 4. Lane 5, site 4 NTC. Lane 6, site 5. Lane 7, site 5 NTC. Lane 8, site 6. Lane 9, site 6 NTC. Lane 10, control site. Lane 11, control site NTC.

The first set of primers designed for both site 1 and site 2 did not successfully amplify the target DNA region, and after several unsuccessful attempts to optimize the reaction conditions (results not shown), were re-designed. The newly designed site 1 primers successfully amplified the desired region at a final primer concentration of 400 nM, as shown in Figure 4-5. The second set of primers designed to amplify the site 2 region, however, were not effective (Figure 4-5).
Reactions were carried out using either 200 nM or 400 nM of the re-designed primer sets for sites 1 and 2. The arrow on the left indicates the position of the site 1 product, which is 131 bp in size. Lane 1, 1 Kb Plus DNA ladder. Lane 2, 200 nM site 1 primers. Lane 3, 400 nM site 1 primers. Lane 4, 200 nM site 2 primers. Lane 5, 400 nM site 2 primers. Lane 6, NTC.

Once again, multiple attempts to optimize the reaction conditions for the site 2 primers were made but were unsuccessful (results not shown), resulting in a second re-design of the primers. This third set of primers was successfully able to amplify the site 2 region, as shown in Figure 4-6, at a final primer concentration of 400 nM.

Reactions were carried out with 400 nM of primers in either the presence or absence of 5% DMSO. The arrow indicates the position of the band, which is 200 bp. Lane 1, 1 Kb Plus DNA ladder. Lane 2, 400 nM site 2 primers. Lane 3, 400 nM site 2 primers with 5% DMSO. Lane 4, NTC.
4.3.3 Amplification Efficiency

Given that all of the primer sets were able to successfully amplify their target regions, the next step was to determine the amplification efficiency of each pair. This was accomplished by preparing a standard curve using duplicate samples containing 2, 3, 5, 10, 25, and 50 ng of DNA and plotting the average Ct value of each reaction versus the log of the quantity of DNA. The slope of the graph was used to identify the amplification efficiency of the primer set, which is required for the analysis of the ChIP-qPCR results. The standard curves for each primer set are shown in Figure 4-7, with their amplification efficiencies listed in Table 4-2.

**Table 4-2: Amplification efficiency of each ChIP primer set**

<table>
<thead>
<tr>
<th>Primer Set</th>
<th>Slope (x)</th>
<th>Amplification Efficiency</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>-3.3044</td>
<td>2.0074</td>
<td>0.9719</td>
</tr>
<tr>
<td>Site 2</td>
<td>-3.4361</td>
<td>1.9545</td>
<td>0.9543</td>
</tr>
<tr>
<td>Site 3</td>
<td>-2.8756</td>
<td>2.2266</td>
<td>0.9837</td>
</tr>
<tr>
<td>Site 4</td>
<td>-3.581</td>
<td>1.9022</td>
<td>0.9977</td>
</tr>
<tr>
<td>Site 5</td>
<td>-3.1228</td>
<td>2.0904</td>
<td>0.9941</td>
</tr>
<tr>
<td>Site 6</td>
<td>-3.8423</td>
<td>1.8208</td>
<td>0.991</td>
</tr>
<tr>
<td>Control Site</td>
<td>-3.5064</td>
<td>1.9284</td>
<td>0.982</td>
</tr>
</tbody>
</table>
Figure 4-7: ChIP primer standard curves
The amplification efficiency of each primer set was determined by performing qPCR using several DNA standards and then plotting the mean Ct value of duplicate samples versus the log of the DNA quantity used in the reaction. The efficiency was calculated using the value of the slope obtained from the equation, as shown in Table 4-2.
Figure 4-7 continued
4.4 Chromatin Immunoprecipitation

4.4.1 Cells

The A498 renal cell carcinoma (RCC) cell line was selected for the ChIP experiment as it expresses endogenous PAX8. The cells were seeded in 10 cm² cell culture dishes and grown to a confluence of 90-95%, which took approximately 48 hours.

4.4.2 Optimization of Cell Lysis and Sonication

In order to obtain chromatin fragments in the target range of 200-1,000 bp, optimal lysis and sonication conditions had to be identified. The first buffer to be tested was the NP-40 lysis buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 10% v/v glycerol, 1% NP-40, 2mM EDTA, protease inhibitors). The cells from four 10 cm² cell culture dishes were scraped from the plastic dishes into 1.5 mL of NP-40 lysis buffer and used to test four different sonication conditions. As can be seen in Figure 4-8, the DNA fragments produced using all four conditions just appear as a smear that runs the length of the gel, indicating that they are much larger than desired.
Figure 4-8: Sonication trial – NP-40 lysis buffer
Chromatin sonicated in NP-40 lysis buffer using four conditions different conditions. Lane 1, 1 Kb Plus DNA ladder. Lane 2, 6 x 10 second pulses. Lane 3, 9 x 10 second pulses. Lane 4, 12 x 10 second pulses. Lane 5, 10 x 15 second pulses. All of the pulses were carried out at 25% amplitude.

Next, the NP-40 buffer was tested alongside the SDS lysis buffer (50 mM Tris-HCl, pH 8.1, 1% w/v SDS, 10 mM EDTA, protease inhibitors) in order to determine if the inefficient sonication was possibly due to the buffer. Also, the number and length of the pulses, as well as the amplitude, were increased. Figure 4-9 shows that sonication in the SDS lysis buffer was much more effective than in the NP-40 lysis buffer. Of the cells lysed in SDS lysis buffer, sonication with ten 20-second pulses produced fragments in the 200 – 650 bp range, while thirteen 20-second pulses yielded 100 – 500 bp fragments, and twenty 20-second pulses resulted in fragments ranging in size from 100 – 400 bp. The cells lysed in NP-40 lysis buffer, on the other hand, produced chromatin fragments ranging in size from 300 – 5,000 bp.
Figure 4-9: Sonication trial – NP-40 lysis buffer versus SDS lysis buffer

Chromatin sonicated in either SDS or NP-40 lysis buffer, using three different conditions. Lane 1, 1 Kb Plus DNA ladder. Lane 2, 10 x 20 second pulses in SDS buffer. Lane 3, 13 x 20 second pulses in SDS buffer. Lane 4, 20 x 20 second pulses in SDS buffer. Lane 5, 10 x 20 second pulses in NP-40 buffer. Lane 6, 13 x 20 second pulses in NP-40 buffer. Lane 7, 20 x 20 second pulses in NP-40 buffer. All pulses were carried out at 30% amplitude.

Finally, the SDS lysis buffer was compared with RIPA buffer (50 mM Tris-HCl, pH 8, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, protease inhibitors). As can be seen in Figure 4-10, cells lysed in SDS lysis buffer produced much more efficiently sonicated chromatin than those that were lysed in RIPA buffer. The chromatin fragments of the SDS-lysed cells ranged in size from <100 – 850 bp with five 20-second pulses, <100 – 600 bp with ten 20-second pulses, and <100 – 450 bp with fifteen 20-second pulses.
Figure 4-10: Sonication trial – SDS lysis buffer versus RIPA buffer

Chromatin sonicated in either RIPA buffer or SDS lysis buffer, under three different conditions. Lane 1, 1 Kb Plus DNA ladder. Lane 2, 5 x 20 second pulses in SDS buffer. Lane 3, 10 x 20 second pulses in SDS buffer. Lane 4, 15 x 20 second pulses in SDS buffer. Lane 5, 5 x 20 second pulses in RIPA buffer. Lane 6, 10 x 20 second pulses in RIPA buffer. Lane 7, 15 x 20 second pulses in RIPA buffer. All pulses were carried out at 25% amplitude.

With the SDS lysis buffer being identified as the optimal buffer for chromatin sonication, the last step was to determine the optimal sonication conditions. Six conditions were tested: five 20-second pulses at 25% amplitude, ten 20-second pulses at 25% amplitude, five 20-second pulses at 30% amplitude, ten 20-second pulses at 30% amplitude, ten 15-second pulses at 25% amplitude and fifteen 15-second pulses at 25% amplitude. Of these, three of the samples (both of those at 30% amplitude and one of the ten 15-second pulses at 25% amplitude) were lost due to the production of bubbles. The remaining three samples were run on a gel and are shown in Figure 4-11. The chromatin fragments produced by the fifteen 15-second pulses ranged in size between 100 – 650 bp, which is smaller than the desired target range. Those produced by the ten 20-second pulses were in the range of 200 – 600 bp range, which is also slightly smaller than desired. The fragments from the five 20-second pulses, on the other hand, were closest to the target range, with sizes of 200 – 850 bp.
Figure 4-11: Sonication trial – number and length of pulses

Chromatin sonicated in SDS lysis buffer using three different conditions. Lane 1, 1 Kb Plus DNA ladder. Lane 2, 5 x 20 second pulses. Lane 3, 10 x 20 second pulses. Lane 4, 15 x 15 second pulses. All pulses were carried out at 25% amplitude.

In order to confirm that this was the optimal setting for sonication and that the results were reproducible, the sonication was repeated with five, seven, eight, and ten 20-second pulses at 25% amplitude. Also, the concentration of the sonicated samples was determined using the NanoDrop prior to running the gel in order to ensure that the same amount of DNA was loaded for each sample. The results of this sonication test are shown in Figure 4-12.
Figure 4-12: Sonication trial – number of 20-second pulses
Chromatin sonicated in SDS lysis buffer using four different conditions. Lane 1, 1 Kb Plus DNA ladder. Lane 2, 5 x 20 second pulses. Lane 3, 7 x 20 second pulses. Lane 4, 8 x 20 second pulses. Lane 5, 10 x 20 second pulses. All pulses were performed at 25% amplitude.

Based on the fact that the five 20-second pulses produced fragments in the 200 – 850 bp range as before and that the remaining samples, with the exception of the seven 20-second pulses, also produced fragments of a similar, although slightly smaller, size, the optimal sonication conditions were determined to be five 20-second pulses at 25% amplitude.

For the actual ChIP experiments, the number of plates of cells used was increased from four to eight. The increased number of cells required an increase in the sonication conditions to ten 15-second pulses, still at 25% amplitude. The efficiency of the sonication was verified prior to continuing to the next step by pooling 1 μL from each aliquot and adding 6 μL of TE buffer, 0.5 μL of proteinase K (10 mg/mL) and 1 μL of RNase A (10 mg/mL). The solution was then incubated at 37°C for 1.5 hours at 1,000 rpm to reverse the crosslinks. The pooled sample was run on a 1.5% agarose gel to ensure fragments of the correct size were obtained, as shown in Figure 4-13.
Figure 4-13: ChIP sonication
Small aliquots of each sonicated chromatin sample obtained during the ChIP experiments were pooled, the crosslinks reversed and the fragments analyzed by agarose gel electrophoresis to ensure that they were in the desired size range of 200 – 1,000 bp. Lane 1, 1 Kb Plus DNA ladder. Lane 2, pooled sonicated chromatin sample.

4.5 ChIP-qPCR
In order to determine to which site(s) PAX8 was binding, the eluted chromatin was analyzed by qPCR using the previously described ChIP primers following incubation with the antibody and immunoprecipitation. Upon examination of the site 2 primer results, the Ct values were quite high, ranging between 31-41. The dissociation curves were also of concern as they showed multiple peaks, indicating the presence of non-specific products (Figure 4-14). While the site 3 results show a single peak corresponding to a single product with nothing in the NTC (non-template control), the site 2 results contain multiple peaks, including in the NTC, indicating several non-specific PCR products and potential contamination of the NTC sample. The qPCR products were therefore run on a 1.5% agarose gel for verification, as shown in Figure 4-15. As can be seen, none of the samples contain a band of the correct size thereby indicating that the qPCR with the site 2 primers was not successful. Due to the numerous problems with the site 2 primers, the qPCR results for this predicted site were not included in the ChIP analysis.
Figure 4-14: Dissociation curves for the site 2 and site 3 qPCR products

The dissociation curves for the no antibody (noAB), PAX8, and input DNA as well as a non-template control (NTC) using primers for site 2 and site 3. The derivative value on the y-axis represents the derivative of the fluorescence as a function of temperature corresponding to the drop in fluorescence that occurs when the SYBR Green dissociates from the DNA as it denatures from double-stranded into single-stranded.
The qPCR products of the no-antibody (noAB), PAX8, and input samples amplified with the site 2 primers were run on a 1.5% agarose gel in order to verify the presence of bands of the correct size. Lane 1, 1 Kb Plus DNA ladder. Lanes 2 and 3, noAB samples. Lanes 4 and 5, PAX8 samples. Lanes 6 and 7, input samples.

The Ct values obtained for each sample are listed in Table 4-3. The average of the duplicate sample values was determined and used in the calculation of the assay site IP fold enrichment.

The dilution factor (DF) between the input DNA and the ChIP DNA (PAX8 and noAB) was 4.5. This was determined by the fact that the input DNA was eluted using 400 µL of 10% Chelex-100 and 50 µL of sterile water (450 µL total) whereas the ChIP DNA was eluted from the beads using 50 µL of 10% Chelex-100 and 50 µL of sterile water (100 µL total).

The DF was then used in combination with the primer amplification efficiency to calculate the adjusted input Ct using the following formula:

\[
\text{adjusted input Ct} = \text{average input Ct} - \log(\text{DF})/\log(\text{primer amplification efficiency})
\]

\[
= 27.20 - \log(4.5)/\log(2.0074)
\]

\[
= 25.0416
\]
Table 4-3: Ct values obtained following each of four biological replicate qPCR experiments amplifying input, PAX8 and no-antibody (noAB) control DNA obtained from ChIP using primers for sites 1, 3-6 and control.

<table>
<thead>
<tr>
<th>Target</th>
<th>Experiment</th>
<th>Input</th>
<th>PAX8</th>
<th>noAB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Site 1</td>
<td>1</td>
<td>27.24</td>
<td>27.16</td>
<td>31.30</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>28.68</td>
<td>28.74</td>
<td>33.00</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>26.76</td>
<td>26.86</td>
<td>31.00</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>28.00</td>
<td>27.70</td>
<td>30.91</td>
</tr>
<tr>
<td>Site 3</td>
<td>1</td>
<td>24.95</td>
<td>25.00</td>
<td>30.73</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>25.68</td>
<td>25.56</td>
<td>30.85</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>24.51</td>
<td>24.51</td>
<td>31.16</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>25.78</td>
<td>25.74</td>
<td>29.30</td>
</tr>
<tr>
<td>Site 4</td>
<td>1</td>
<td>25.06</td>
<td>25.00</td>
<td>30.64</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>25.85</td>
<td>26.00</td>
<td>31.87</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>24.80</td>
<td>24.91</td>
<td>27.00</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>25.35</td>
<td>25.41</td>
<td>29.00</td>
</tr>
<tr>
<td>Site 5</td>
<td>1</td>
<td>25.00</td>
<td>24.77</td>
<td>30.47</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>24.75</td>
<td>25.00</td>
<td>30.35</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>24.15</td>
<td>23.94</td>
<td>26.00</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>24.90</td>
<td>25.00</td>
<td>28.72</td>
</tr>
<tr>
<td>Site 6</td>
<td>1</td>
<td>24.23</td>
<td>24.22</td>
<td>30.09</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>24.96</td>
<td>25.00</td>
<td>30.79</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>23.72</td>
<td>24.00</td>
<td>25.70</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>24.50</td>
<td>24.41</td>
<td>28.13</td>
</tr>
<tr>
<td>Control Site</td>
<td>1</td>
<td>24.18</td>
<td>24.05</td>
<td>29.62</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>25.00</td>
<td>24.78</td>
<td>30.95</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>24.08</td>
<td>24.17</td>
<td>31.21</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>24.77</td>
<td>25.00</td>
<td>30.82</td>
</tr>
</tbody>
</table>
The adjusted input Ct value was subtracted from the ChIP Ct values to calculate the input normalized IP value (ΔCt):

$$\Delta C_t\ (PAX8) = \text{average } C_t - \text{adjusted input } C_t$$
$$= 31.48 - 25.0416$$
$$= 6.44$$

$$\Delta C_t\ (\text{noAB}) = 33.72 - 25.0416$$
$$= 8.68$$

These values were then used to determine the noAB adjusted IP Ct value ($\Delta\Delta C_t$), which represents the difference in Ct between the PAX and noAB IP signals:

$$\Delta\Delta C_t = \Delta C_t\ (PAX8) - \Delta C_t\ (\text{noAB})$$
$$= 6.44 - 8.68$$
$$= -2.24$$

Finally, this value was used to calculate the assay site IP fold enrichment (above background):

$$2^{-(\Delta\Delta C_t)} = 2^{-(2.24)}$$
$$= 4.8$$

The assay site IP fold enrichment values for each primer set are listed in Table 4-4. The mean of the values was calculated and then divided by that of the control site to determine the fold enrichment of each predicted binding site relative to the control site (Figure 4-16).
Table 4-4: Assay site IP fold enrichment values for each predicted binding site

<table>
<thead>
<tr>
<th>Target</th>
<th>Experiment</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>4.80</td>
<td>1.70</td>
<td>0.20</td>
<td>0.90</td>
</tr>
<tr>
<td>Site 3</td>
<td>9.50</td>
<td>27.90</td>
<td>8.60</td>
<td>1.30</td>
</tr>
<tr>
<td>Site 4</td>
<td>8.50</td>
<td>7.90</td>
<td>0.90</td>
<td>0.70</td>
</tr>
<tr>
<td>Site 5</td>
<td>12.00</td>
<td>14.40</td>
<td>1.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Site 6</td>
<td>16.00</td>
<td>10.30</td>
<td>1.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Control Site</td>
<td>22.90</td>
<td>9.60</td>
<td>6.00</td>
<td>4.90</td>
</tr>
</tbody>
</table>

Figure 4-16: Fold change of each predicted binding site relative to the control site

The mean fold enrichment of each predicted PAX8 binding site was divided by that of the control site in order to determine the fold change and thus the most likely candidate(s) for PAX8 binding in the region of the E2F1 promoter 728 bp upstream of the transcription start site. The values shown represent the mean +/- 1 SEM.
Site 3 is the sole predicted site with a fold enrichment higher than that of the control site, albeit only 1.25-fold (P > 0.05). This suggests that PAX8 may be binding to site 3 in the E2F1 promoter. The fact that the fold enrichment values for the control site are consistently higher than those of the other predicted binding sites (with the exception of site 3) suggests that while it was not predicted to bind PAX8, that may not be the case. Site 1, on the other hand, has the lowest fold enrichment values of all the predicted sites indicating that PAX8 may not bind at this location. Taken together, this information suggests that site 1 might be a better control site to use in determining the fold changes of the other sites as the results suggest that it is more likely to not be binding PAX8 than the control site that was selected. Using site 1 to calibrate the results suggests an even greater likelihood that PAX8 is binding to site 3 with an enrichment of 15.71-fold (P > 0.5) relative to site 1 (Figure 4-17). It is also likely that, while not in the region of interest, PAX8 may bind to the control site as it has a 11.47-fold (P > 0.5) greater enrichment as compared to site 1.

Figure 4-17: Fold enrichment of each predicted PAX8 binding site relative to site 1

The mean fold enrichment of each predicted PAX8 binding site was divided by that of site 1, which is suggested to be more likely to not bind PAX8 than the original site that was chosen for use as the control, in order to determine the fold change and thus the most likely candidate(s) for PAX8 binding in the region of the E2F1 promoter 728 bp upstream of the transcription start site. The control site is further upstream of the region of interest. Values are mean +/- 1 SEM.
4.5.1 Amplicon Sequencing

In order to confirm that the primers were amplifying the correct target sequences, the qPCR products for each primer set were sequenced. The products were first run on a 1.5% agarose gel to confirm the presence of only a single band (Figure 4-18). The bands were then excised and the DNA extracted from the gel. Cycle sequencing was carried out with the purified DNA, and then the samples were cleaned up and sent to the Anatomy Department for dideoxy sequencing. The alignment of each sequenced qPCR product with E2F1 reference sequence is shown in Figure 4-19.

![ChIP-qPCR product size range](image)

**Figure 4-18:** ChIP-qPCR products to be sequenced

The ChIP-qPCR products for each predicted binding site were run on a 1.5% agarose gel to ensure the presence of a single band of the correct size in each sample. The bands were then excised from the gel, the DNA purified and sequenced. Lane 1, 1 Kb Plus DNA ladder. Lane 2, site 1. Lane 3, site 3. Lane 4, site 4. Lane 5, site 5. Lane 6, site 6. Lane 7, control site.
Figure 4-19: Sequence alignment of ChIP-qPCR products with the \textit{E2F1} reference sequence

The sequences obtained from the ChIP-qPCR products of each of the predicted PAX8 binding sites were aligned with the \textit{E2F1} reference sequence using ClustalW2 to determine if the correct region had been amplified. The locations bound by the primers are shown in bold and the target sites are highlighted in green. Asterisks (*) indicate exact matches and dashes (-) indicate gaps in the sequence either due to missing bases or inserted by ClustalW2 to obtain better alignment.
4.6 Discussion

4.6.1 ChIP

The ChIP results suggest that site 1 does not bind PAX8 and therefore it was decided to use this site as the control as opposed to the control site that was originally selected, which in Figure 4-16, appears to bind PAX8. Site 3 also appears to bind PAX8 with a fold enrichment of 15.7-fold (P > 0.5) greater than...
that of site 1. While there is a large variation in the results for each IP assay site with respect to the fold enrichment values obtained in each experiment (Table 4-4), a trend is observed in the results of experiments 2-4 whereby site 3 has the highest fold enrichment amongst the predicted PAX8 binding sites. The large amount of variation and the fact that the control site showed enrichment are possible contributing factors to the lack of statistical significance of the results. This, in combination with the absence of site 2 from the experiment, does not allow any definitive conclusions to be made as to which specific site(s) PAX8 is binding.

### 4.6.2 ChIP-qPCR and Sequencing

The ChIP-qPCR products for site 5 and the control site were successfully sequenced over the entire length of the amplicon thereby demonstrating that the desired target region was amplified. The remaining sequences all contained mismatches and gaps, which would not normally be expected for sequencing on this scale. These issues, combined with the short read lengths that were obtained (for chromatograms see Appendix C) could be the result of a variety of different factors including too much or too little template DNA, improperly purified template DNA, excessive dilution of the BigDye reagent, or too much primer. Although the sequenced products for sites 1, 4, and 6 contain some mismatches and gaps, the actual predicted PAX8 binding sites were successfully sequenced suggesting that the correct region was amplified. The site 3 amplicon also contains mismatches and gaps including in the predicted binding site region. Multiple attempts to sequence this amplicon were made but unfortunately none of those using the forward primer extended far enough to include the binding site. As for the sequencing reactions using the reverse primer, a complete sequence for the binding site was not obtained possibly due to the fact that the binding site was immediately 5’ of where the primer bound. Despite these issues, the fact that the region upstream of the binding site and the majority of the binding site were successfully sequenced suggests that the correct region was indeed amplified.
4.6.3 Predicted Binding Sites

Several PAX8 consensus binding sequences have been published, including: MTGCCCASTCAAGYKT in the $Tg$ and $TPO$ promoters (Zannini et al., 1992), NNTNNYGCNGTARGAR from a pool of 64-bp oligonucleotides (Pellizzari et al., 1996), CASTSANGCNK in the $WT1$ promoter (Fraizer et al., 1997), and TNNNGCGKRAVSR in the $NIS$ upstream enhancer (Ohno et al., 1999), where $M = A$ or $C$, $S = C$ or $G$, $Y = C$ or $T$, $K = G$ or $T$, $R = A$ or $G$, and $N = \text{any base}$. Jun and Desplan (1996) published a PAX8 consensus binding sequence representing the optimal PAX8 binding site: NMNGTCACSCNYNASTR. The sequence used by ConTra in the V$PAX8_01$ matrix to identify sites predicted to bind PAX8 in the $E2F1$ promoter was a consensus sequence from the TRANSFAC database – NNNTNNGNGTGANN (Matys et al., 2003).

The TRANSFAC sequence contains only six invariant bases with the rest being completely variable. When aligned with the predicted PAX8 binding sites (Figure 4-20), it can be seen that only site 4 contains all six invariant bases, and yet, the ChIP results indicated that this particular site is not binding PAX8 with any great affinity. Site 3, on the other hand, contains only three of the six invariant bases in the TRANSFAC sequence and the ChIP results suggested that this is the site that is binding PAX8 with the greatest affinity.
**TRANSFAC**

```
NNN\_NNN\_NNN\_G\_TGA\_NNN\-
```

**Site 1**

```
-CTCACGG-CCGGGC
```

**Site 2**

```
-ATTCACCCCGCCT-
```

**Site 3**

```
-GGTACCT-GCAGAG
```

**Site 4**

```
ATGTCACTG-GTGAAG-
```

**Site 5**

```
GGGCTCTGG-GTGAAG-
```

**Site 6**

```
--TGAATGA-GTGAAGGG
```

**Figure 4-20:** Alignment of the TRANSFAC consensus with the predicted PAX8 binding sites.

The PAX8 consensus sequence from the TRANSFAC database used by the VSPAX8_01 position weight matrix in ConTra to predict PAX8 binding sites in the *E2F1* promoter is aligned with the six predicted binding sites. The bases highlighted in blue in the TRANSFAC sequence represent the six invariant bases, with the matches to these bases highlighted in blue in the predicted site sequences. Dashes represent gaps inserted into the sequences to obtain the best possible alignment.

The optimal PAX8 binding consensus sequence identified by Jun and Desplan (1996) is more invariant than the TRANSFAC consensus sequence (Figure 4-21) and provides some greater insight into the ChIP results. They identified a core motif consisting of seven bases, which are conserved between the optimal binding sites of the paired domains of Prd, Pax8, Pax2, Pax5, and Pax6. None of the predicted sites contain all seven of these conserved bases, although sites 1, 2, and 3 all contain six and site 4 has five (Figure 4-22). It is interesting to note that while site 1 and site 3 share the same six conserved core motif bases, site 3 binds to PAX8 while site 1 does not. Reasons for this may include the identity of the seventh core motif base, a G in site 1 versus a T in site 3, and/or the identity of the remaining bases in the binding site sequences, and/or sequences surrounding the predicted binding sites.
Figure 4-21: PAX8 consensus sequence alignment
Alignment of the TRANSFAC consensus sequence used to identify predicted PAX8 binding sites in the E2F1 promoter and the optimal PAX8 binding site consensus sequence identified by Jun and Desplan (1996). M = A or C; S = G or C; Y = T or C. Exact matches are highlighted in blue while matches to a base with two possible identities are highlighted in yellow.

Figure 4-22: Alignment of the optimal PAX8 binding site consensus sequence with the predicted PAX8 binding sites.
The optimal PAX8 binding site consensus sequence identified by Jun and Desplan (1996) is aligned with the predicted PAX8 binding sites identified with the TRANSFAC consensus sequence. The seven bases that make up the core motif are numbered and boxed. M = A or C; S = G or C; Y = T or C. Exact matches are highlighted in blue while matches to a base with two possible identities are highlighted in yellow.

When comparing the alignments of the predicted sequences with the TRANSFAC and optimal PAX8 binding site sequences, it is interesting that sites 5 and 6, which show the greatest sequence similarity to the TRANSFAC sequence (Figure 4-20), show the least similarity to the optimal PAX8 binding site consensus sequence (4-22). The fact that neither of these sites appears to bind PAX8 with any great affinity suggests that the optimal PAX8 binding site consensus sequence identified by Jun and Desplan (1996) is a better predictor of functional binding sites. That being said, this theory does not work for sites 1 and 4, which have the same number of matching bases as site 3, but have the lowest abundances in the ChIP-qPCR results (Figure 4-16). As Jun and
Desplan (1996) noted, while the consensus sequence they identified represents the optimal binding site sequence, not all functional binding sites will have the optimal sequence. The reverse also likely holds true in that not all predicted binding sites having the optimal sequence are functional, as is the case with site 1 and site 4.

PAX8 binding sites have been identified in several other promoters, including those of the thyroglobulin and thyroperoxidase genes (Tg and TPO; Zannini et al., 1992) as well as the Wilms’ tumour gene (WT1; Fraizer et al., 1997). When the predicted sequences are aligned with the known PAX8 binding sites (Figure 4-23), it can been seen that there is extensive sequence homology between the three published sites, as well as between the core motifs of the predicted sites and the published sites. Of note is the fact that site 3, identified as binding PAX8 in the E2F1 promoter, has seven homologous bases, which is higher than all of the other predicted sites save for site 2. Site 2 has 8 homologous bases, suggesting that perhaps it too binds PAX8 since it shares the most homologous bases with three previously identified PAX8 binding sites.

**Figure 4-23**: Alignment of the predicted binding sites with the published PAX8 binding sites

The six predicted binding sites identified in the E2F1 promoter are aligned with the published PAX8 binding sites in three other promoters: thyroglobulin (Tg; Zannini et al., 1992), thyroperoxidase (TPO; Zannini et al., 1992), and Wilms’ tumour gene (WT1; Fraizer et al., 1997). The bases highlighted in blue indicate matches between sequences, with those highlighted in yellow indicating a second set of matching bases at the same position.
4.6.4 Site 2

4.6.4.1 qPCR

The question of whether or not PAX8 binds to the predicted site 2 sequence could not be answered as a set of primers to amplify this region was unable to be designed for the ChIP-qPCR. While the third set of primers designed for this site yielded a product of the correct size during the reaction optimization process, this result could not be replicated when the reaction was carried out on the qPCR machine. There are several potential contributing factors as to why the ChIP-qPCR was unsuccessful for the site 2 primers including the target region of the E2F1 promoter containing the predicted site, the primers themselves, and the reaction conditions.

Amplicon

The amplicon size of the site 2 region was 197 bp. This puts it at the upper limit of the target length of 60-200 bp for qPCR. It is also much longer than the other amplicons, which range in size from 131-150 bp. A relatively short target size is ideal for qPCR as longer products do not amplify as efficiently as shorter ones.

Another potential issue with this particular region of the E2F1 promoter is its very high GC content of 75%, while the ideal GC content for qPCR amplicons is 50%. The other predicted binding sites are located in areas with lower GC contents, ranging from 47-63%. GC-rich templates can be difficult to amplify due to the possible formation of secondary structures, such as hairpins, and higher melting temperatures. Along with the high GC content, there are also significant stretches of G’s and C’s in this region, which can cause the polymerase to slip during the amplification process.

Primer Design

During the reaction optimization process, the site 2 primers successfully amplified the target region at a primer concentration of 400 nM. This reaction was tested with and without the addition of 5% DMSO, which is often used
with GC-rich template DNA to inhibit the formation of secondary structures. When run on a 1.5 % agarose gel, the product of the reaction with the DMSO was observed to stain with less intensity as compared to the product from the reaction without the DMSO. Primer dimers were also present in both products. Given the fact that the qPCR using these primers was not successful and amplified two products, titration of the primer concentration could be done to eliminate the presence of the primer dimers.

Upon examining the primers themselves, they fall into the ideal range of 18-24 nucleotides in length, with 21 nt for the forward primer and 20 nt for the reverse. With a GC content of 61.9% for the forward and 60.0% for the reverse, these values are slightly higher than the approximately 50% that is desired in order to assist in the prevention of mismatch stabilization. Both primers also have stretches of G’s and C’s, with the reverse primer containing a CGG triplet. While the forward primer contains a GC clamp at the 3’ end, the reverse primer does not, and this is another feature that is desirable in qPCR primers. All of this information indicates potential issues with the primers, particularly the reverse, and thus provides further evidence as to why the site 2 qPCR was not successful.

**Reaction Conditions**

The success of the PCR as compared to the qPCR suggests that some component that was different between the two reactions may have contributed to the fact that the target amplicon was not successfully amplified by qPCR. The two reactions contained the same amount of template DNA as well as equal concentrations of all of the other components. Both reactions also started out being heated at 50°C for 2 minutes and then 95°C for 10 minutes. This was followed by 45 cycles, although the temperature settings and times of the cycles differed between the two reactions. The PCR cycles consisted of 95°C for 30 seconds, 60°C for 1 minute, and 72°C for 1 minute, whereas the qPCR cycle settings were 95°C for 15 seconds and 60°C for 1 minute. The PCR finished with a final extension step at 72°C for 7 minutes. The qPCR, on the other hand, had an additional dissociation step of 95°C for 15 seconds, 60°C for 1 minute,
and 95°C for 15 seconds. Based on this comparison, the most noticeable difference between the two reactions is the absence of the 72°C extension steps in the qPCR. Given that the amplicon is quite long for qPCR at 197 bp, it is possible that the absence of the extension step resulted in an inefficiently amplified target sequence.

Given the length of the target, its high GC content, the occurrence of long stretches, of G’s and C’s, and the presence of primer dimers, it is perhaps understandable as to why the qPCR was unsuccessful. While the amplicon length and primer concentration can be changed, the GC content cannot, which is why three sets of primers were designed for this region without any success and why successful qPCR will be difficult to achieve. A recommendation for future attempts to design primers for this region would involve using a shorter amplicon length and to work at optimizing the reaction with the addition of DMSO in an attempt to counteract secondary structure formation.

**4.6.4.2 PAX8 Binding**

When compared with the optimal PAX8 binding site consensus sequence (Figure 4-22), the site 2 sequence contains all six of the fixed bases in the core motif as well as an additional match outside the core motif. Finally, when aligned with the three published PAX8 binding sites from the Tg, TPO, and WTI promoters, site 2 had the greatest homology out of all of the predicted binding sites, with eight matching bases (Figure 4-23). In the previous chapter, the results of the promoter-reporter assay suggest the presence of two PAX8 binding sites in the 728 bp region upstream of the transcription start site in the E2F1 promoter, with one being in the 242 bp region upstream of the start site. All of this combined with the fact that the ChIP results indicate that site 1, the only other predicted binding site in the 242 bp region upstream of the transcription start site, does not bind PAX8, suggest that PAX8 binds to site 2.

Further support of the predicted binding of PAX8 to site 2 comes from its homology to the optimal binding site sequence. As shown in Figure 4-24, there
is a high homology between the Pax8 and the Drosophila paired protein (Prd) consensus binding sequence, particularly in the core motif. When the base in the seventh position of the core motif of Prd was mutated, Jun and Desplan (1996) found that there was a resulting 75-fold decrease in the binding affinity of its paired domain (PrdPD). Xu and colleagues (1995) crystallized the PrdPD bound to an optimized binding site in a 15 bp fragment of DNA and solved the structure. When aligned with the sequence of the PAX8 paired domain (PAX8PD; Poleev et al., 1992), it can be seen that all of the amino acids that contact the DNA in the PrdPD are conserved in the PAX8PD (Figure 4-25). Since the seventh base in the core motif forms a direct contact with Prd via a hydrogen bond with the glycine at position 15 (Xu et al., 1995) and this amino acid is conserved in the PAX8PD, it is reasonable to think that this amino acid is also forming a direct contact with the seventh base in the core motif of the PAX8 binding site consensus sequence. Site 2 is the only one of the predicted sites that contains the same base at position 7 as the optimal binding site sequence, suggesting that this might increase its affinity for PAX8. However, it is interesting to note from Figure 4-1 that neither site 2 nor site 3 are very strongly conserved across a wide range of species, which perhaps may relate to a species-specific role in the regulation of E2F1 by PAX8.

**Figure 4-24:** The optimal binding site sequences for Pax8 and the Drosophila paired (Prd) protein

Alignment of the Pax8 and Prd optimal binding site sequences show significant homology. Exact matches are highlighted in blue while matches to a base with two possible identities are highlighted in yellow.
Figure 4-25: The secondary structure and amino acid sequence of the Drosophila paired protein paired domain (prd; Xu et al., 1995) and the amino acid sequence of the PAX8 paired domain (PAX8; Poleev et al., 1992).

Asterisks indicate matching amino acids between the two sequences. The prd amino acids involved in forming contacts with DNA are shown below the sequence. The top line indicates those that contact the sugar phosphate backbone (p), while the bottom line indicates those that contact bases in either the major groove (M) or the minor groove (m). The amino acid highlighted in blue is the glycine at position 15 that forms a contact with the seventh base in the core motif of the optimal PAX8 consensus binding site identified by Jun and Desplan (1996). Adapted from Xu et al., 1995.
CHAPTER 5
– PAX8 Binding Site Validation –

5.1 Introduction
The results presented in the previous two chapters suggest that PAX8 activates the $E2F1$ promoter and that it binds directly to site 3. Unfortunately, it was not possible to determine if PAX8 binds to site 2 as primers to amplify this G-C rich region were not able to be successfully designed for the ChIP experiment. In order to validate the results of the ChIP experiment, as well as to determine if PAX8 may also be binding to site 2, two experiments were carried out.

First, deletion constructs were created in order to sequentially eliminate each of the predicted binding sites from the $E2F1$ promoter and then the activity of each construct was tested using the luciferase assay. The goal of this experiment was to determine if PAX8 may also be binding to site 2, and to examine the effects of each individual predicted binding site on the activation of the $E2F1$ promoter.

The second experiment involved the use of site-directed mutagenesis. Three bases in the site 3 binding site sequence were mutated and then the activity of the mutant construct was tested and compared to that of the pE2F1(-242) and pE2F1(-728) constructs in order to confirm that PAX8 is activating the $E2F1$ promoter by binding to site 3.

5.2 Deletion Constructs
The first step in the validation process was to create deletion constructs in order to determine the effects of each binding site on the activation of the $E2F1$ promoter. Site 2 was of particular interest as it was not able to be included in the ChIP experiment.
5.2.1 Cloning

5.2.1.1 Primers

Primers were designed in order to amplify particular regions of the \textit{E2F1} promoter, using the pE2F1(-728) construct as the template, as outlined in Figure 5-1. The forward primers were all designed to incorporate the NheI restriction enzyme recognition sequence and the reverse primer was designed to incorporate the HindIII restriction enzyme recognition sequence for cloning purposes (Figure 5-2).

\textbf{Figure 5-1: Deletion Constructs}

The deletion constructs were designed to sequentially delete each predicted PAX8 binding site located in the 728 bp region upstream of the transcription start site (position +1) of the \textit{E2F1} promoter.
Figure 5-2: Deletion Construct

The inserts for each deletion construct (including F1, as shown) were cloned from the pE2F1(-728) construct using primers that incorporated the NheI and HindIII restriction enzyme recognition sites. These restriction enzymes were used to clone the inserts into the pGL2-basic vector upstream of the luciferase gene (luc).

5.2.1.2 Amplification

The target insert sequences were amplified using the proofreading enzyme Phusion Hot Start II DNA Polymerase (Finnzymes, Finland) in order to reduce the occurrence of errors during PCR. DMSO was also added to a final concentration of 5% in all of the reactions as the E2F1 promoter region used as the template is very G-C rich. The PCR products obtained from each reaction are shown in Figure 5-3. A sample using the same template DNA and the GL3F2/R1 control primers was included as a positive control.
Figure 5-3: Insert DNA for deletion constructs

PCR products of amplified DNA from the pE2F1(-728) construct. Lane 1, 1 Kb Plus DNA ladder. Lane 2, positive control. Lane 3, F1 insert (683 bp). Lane 4, F2 insert (600 bp). Lane 5, F3 insert (483 bp). Lane 6, F4 insert (423 bp). Lane 7, F5 insert (197 bp). Lane 8, F6 insert (122 bp).

5.2.1.3 DNA Purification

As can be seen in Figure 5-3, the PCR products for the F1, F2, F3, F4, and F5 inserts contained multiple bands. In order to purify the insert DNA, the remainder of the PCR products for these five inserts was run on a second agarose gel, the desired bands were excised, and the DNA was purified from the agarose. As the F6 PCR product was a clean, single band, the PCR produce was purified directly from the PCR reaction.

5.2.1.4 Restriction Digests

The purified insert DNA and the pGL2-basic vector were then digested with both NheI and HindIII. Following digestion, the DNA was run on an agarose gel for verification, shown in Figure 5-4. Once again the remainder of the successfully digested DNA was run on a second agarose and purified in preparation for the ligation step.
DNA digested in preparation for the ligation of the insert and vector portions to form the deletion constructs. A) Lane 1, 1 Kb Plus DNA ladder. Lane 2, undigested pGL2-basic. Lane 3, HindIII-digested pGL2-basic. Lane 4, NheI-digested pGL2-basic. Lane 5, HindIII/NheI-digested pGL2-basic. Lane 6, HindIII/NheI-digested F5 insert. Lane 7, HindIII/NheI-digested insert. The bracket indicates the locations of the digested insert DNA bands. B) Lane 1, 1 Kb Plus DNA ladder. Lane 2, HindIII/NheI-digested pGL2-basic. Lane 3, HindIII/NheI-digested F1 insert. Lane 4, HindIII/NheI-digested F2 insert. Lane 5, HindIII/NheI-digested F3 insert. Lane 6, HindIII/NheI-digested F4 insert.

5.2.1.5 Ligation and Transformation

Insert and vector DNA were ligated together using T4 DNA ligase. The ratio used was 3:1 insert:vector, with 90 fmol of insert ends and 30 fmol of vector
The quantities of insert and vector DNA were calculated using the following equations:

Vector DNA = 30 fmol x 1 µg/3000 fmol x size of vector (bp)/1000 bp

\[ = 30 \times \frac{1}{3000} \times \frac{5600}{1000} \]
\[ = 0.056 \mu g \]
\[ = 56 \text{ ng} \]

Digested pGL2-basic = 23.56 ng/µL

\[ \therefore \text{ Vector DNA } = \frac{56 \text{ ng}}{23.56 \text{ ng/µL}} \]
\[ = 2.4 \mu L \]

F6 insert DNA = 90 fmol x 1 µg/3000 fmol x size of insert (bp)/1000 bp

\[ = 90 \times \frac{1}{3000} \times \frac{122}{1000} \]
\[ = 0.004 \mu g \]
\[ = 4 \text{ ng} \]

Digested F6 = 6.09 ng/µL

\[ \therefore \text{ F6 insert DNA } = \frac{4 \text{ ng}}{6.09 \text{ ng/µL}} \]
\[ = 0.7 \mu L \]

The F5 and F6 inserts were used to test two different ligation conditions: 14°C for 24 hours and room temperature for 4 hours. The newly ligated DNA was then transformed into competent DH5α cells and plated. Following an overnight incubation at 37°C, the plates with DNA that was ligated at 14°C for 24 hours had a greater number of colonies than those with DNA that was ligated at room temperature for 4 hours (Table 5-1), therefore the remaining four inserts were ligated to the pGL2-basic vector for 24 hours at 14°C.
### Table 5-1: Number of colonies obtained from transformations done following ligation of the F5 and F6 inserts to the pGL2-basic vector using two different ligation conditions

<table>
<thead>
<tr>
<th>Insert</th>
<th>Ligation Conditions</th>
<th>Number of Colonies</th>
</tr>
</thead>
<tbody>
<tr>
<td>F5</td>
<td>14°C for 24 hours</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>RT for 4 hours</td>
<td>7</td>
</tr>
<tr>
<td>F6</td>
<td>14°C for 24 hours</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>RT for 4 hours</td>
<td>12</td>
</tr>
</tbody>
</table>

#### 5.2.1.6 Single Colony PCR

Multiple colonies were selected from each plate and tested to see if they contained the desired insert. Small samples of each culture were incubated at 94°C for 2 minutes and then placed on ice for a minimum of 5 minutes in order to lyse the cells prior to the addition of the PCR mix. The plasmid DNA in the cultures was then amplified using the corresponding primers and the presence of the correct insert was confirmed by running the PCR products on a 1.5% agarose gel, as shown in Figure 5-5. The top two gels represent the results of the ligation condition trial using the F5 and F6 inserts. As can be seen, all of the colonies selected from the F6 plates contain the desired insert, regardless of which ligation condition was used. The colonies selected from the F5 plates, on the other hand, did not contain the desired insert for either ligation condition. The bands visible on the gel are most likely primer dimers. The other three gels represent the results from the ligation of the remaining four inserts, along with the repeated ligation using the F5 insert. In this instance, three of the eight F1 colonies, three of the eight F2 colonies, three of the five F3 colonies, one of the two F4 colonies, and four of the seven F5 colonies contained the desired insert.
Figure 5-5: Single colony PCR products used to verify presence of insert DNA

a) – e) Lane 1, 1 Kb Plus DNA ladder. A) Lanes 2-14, F6 deletion constructs ligated for 24 hours at 14°C. b) Lanes 2-5, F6 deletion constructs ligated for 4 hours at room temperature. Lanes 6-10, F5 deletion constructs ligated for 24 hours at 14°C. Lanes 11-13, F5 deletion constructs ligated for 4 hours at room temperature. C) Lanes 2-6, F1 deletion constructs. Lane 7-11, F2 deletion constructs. Lanes 12-15, F3 deletion constructs. D) Lanes 2-4, F1 deletion constructs. Lanes 5-7, F2 deletion constructs. Lane 8, F3 deletion construct. Lanes 9-10, F4 deletion constructs. E) Lanes 2-8, F5 deletion constructs.
5.2.1.7 DNA Purification and Sequencing

Plasmid DNA was purified from cultures that were identified to contain the deletion construct with the desired insert using minipreps and then sent to the Anatomy Department for sequencing. The sequencing results for each of the constructs were then aligned with the sequences of the E2F1 promoter and the pE2F1(-728) plasmid in order to confirm that the constructs had been cloned correctly. Sequencing was performed using the GL3F2 and GL3R1 control primers to verify that the insert was in the proper orientation as both of these primers bind to the pGL2-basic vector outside of the multiple cloning site (Figure 5-6). The successfully sequenced deletion constructs are found in Appendix D.

Figure 5-6: pGL2-basic vector depicting the locations of the GL3F2 and GL3R1 primers

The GL3F2 and GL3R1 bind to the pGL2-basic vector outside of the multiple cloning site, as shown by the green arrows. These primers were used in the sequencing reactions to sequence the deletion construct inserts that were cloned into the vector between the NheI and HindIII sites.
Following successful sequencing, the deletion constructs were purified from 100 mL cultures using the Invitrogen Midiprep Kit and quantified using a NanoDrop.

5.2.2 Promoter-Reporter Assay

5.2.2.1 PAX8 Activation

Once purified by midiprep, the activation of the deletion constructs by PAX8 was tested, the results of which are shown in Figure 5-7. The pE2F1(-242) and pE2F1(-728) constructs were included as controls along with pGL2-basic vector.

Figure 5-7: E2F1 promoter construct activity following overexpression of PAX8

The E2F1 promoter constructs [pE2F1(-242) and pE2F1(-728)], deletion constructs, and pGL2-basic vector were co-transfected with either 0 ng or 400 ng of PAX8 expression vector and the resulting luciferase activity was measured. The values shown represent the mean +/- 1 SEM. * P < 0.05 (unpaired two-tailed t test)
As can be seen, the activity levels of the pE2F1(-242) and pE2F1(-728) constructs are much higher as compared with those of the deletion constructs. When looking at the deletion construct results alone (Figure 5-8), it is apparent that the activity levels of the deletion constructs, with the exception of F1, are similar to that of the pGL2-basic vector, which does not contain a promoter. The practically non-existent activity levels of the F4 deletion construct suggest that there may be something wrong with its construction as it contains the same predicted binding sites as the pE2F1(-242) construct with some extra bases upstream.

**Figure 5-8:** The activity levels of the pGL2-basic negative control and the deletion constructs, with (a) and without (b) the data from the longest construct, F1.

The values shown represent the mean +/- 1 SEM.
5.2.2.2 E2F1 Activation

Given the fact that the activity levels of all of the deletion constructs were much lower than expected, the next step was to co-transfect each of the constructs with E2F1, which is known to be the major activator of the E2F1 promoter. As can be seen in Figure 5-9, the same pattern of activity was observed following the co-transfection of the promoter constructs with E2F1. These results were also not as expected given the fact that all of the constructs contain the same published E2F1 binding site just upstream of the transcription start site. Looking at the difference in the activity levels between the pE2F1(-242) and pE2F1(-728) constructs, suggests that as with PAX8 binding sites, there may be more than one site to which E2F1 is binding and activating its promoter. Again, when examining the five shortest constructs, similar results are observed (Figure 5-10) as were seen with PAX8 overexpression.

![Figure 5-9: The activity levels of the various E2F1 promoter constructs co-transfected with E2F1.](image)

The E2F1 promoter constructs, pE2F1(-242) and pE2F1(-728), deletion constructs, and pGL2-basic vector were co-transfected with either 0 ng or 100 ng of E2F1 expression vector and the resulting luciferase activity was measured.
Figure 5-10: The activity levels of the deletion constructs following co-transfection with E2F1, with (a) and without (b) the F1 data.

5.2.3 Deletion Construct Analysis

The fact that the deletion constructs did not perform as expected prompted an examination of the plasmids.
5.2.3.1 Insert Orientation

The deletion constructs were analyzed to ensure that the inserts were cloned in the proper orientation. PCR with the insert-specific forward primers and a reverse primer that bound to the vector were used in order to accomplish this. As can be seen in Figure 5-11, the inserts appear to be present in the correct orientation in all of the constructs except F4. The cloning of F4 in the reverse orientation would account for its practically non-existent activity levels in the promoter-reporter assay. While F5 appears only as a faint band on the gel in comparison to the other inserts, its presence indicates that it was indeed inserted in the proper orientation.

Figure 5-11: Deletion construct insert orientation

The orientation of the inserts in each of the deletion constructs was determined by amplifying the region using forward primers specific to each insert and a reverse primer that binds to the pGL2-basic vector. Lane 1, 1 Kb Plus DNA ladder. Lane 2, F1 insert. Lane 3, F2 insert. Lane 4, F3 insert. Lane 5, F4 insert. Lane 6, F5 insert. Lane 7, F6 insert.

5.2.3.2 Deletion Mutants

As noted in Table 2-2, Prof. Ohtani pointed out that deletion mutants sometimes occur (visible as a smaller fragment following single digest) during expansion of the pE2F1(-728), so each deletion construct was digested with Smal and SacI in order to verify that this was not the case (Figure 5-12).

The Smal-digested and undigested plasmid DNA for all six of the deletion constructs have the same appearance on the gel indicating one of three things:
either the SmaI enzyme was not functional, the digestion reaction was not carried out for a long enough time, or the SmaI recognition sequence was missing from the plasmids, possibly lost during the cloning process. Sequencing of the deletion constructs (Appendix D) indicates that the SmaI recognition site is indeed present and thus the problem most likely had something to do with either the enzyme itself or the reaction conditions. Lengthening of the reaction time or use of a different aliquot of the enzyme would have allowed for clarification of this issue. The SacI enzyme, on the other hand, appears to have partially linearized the F1, F2, F3 and F5 plasmids as indicated by the presence of two bands on the agarose gel. The SacI-digested F6 plasmid is present as a single band, indicating complete digestion. While the SacI-digested F4 plasmid DNA is also present as a single band, it is not of the correct size, suggesting that the plasmid was not successfully digested and may therefore be missing the SacI recognition sequence. The lack of a smaller fragment in the fully digested F6 plasmid indicates that there was not deletion mutant contamination in this sample. In addition, the absence of any smaller fragments present in the other SacI-digested plasmids thereby suggests that they are also not contaminated by deletion mutants, however this could only be said for certain following complete digestion of each of the plasmids as opposed to the partial digestion that is shown.

Undigested plasmid DNA is normally visible on an agarose gel as multiple bands of different sizes, representing supercoiled DNA, nicked open circular DNA, and linear DNA. The undigested F2, F3, and F4 plasmids present as single bands, while F1, F5, and F6 all appear as multiple bands. Given that the plasmids are all made up of the same vector and inserts of only slightly different sizes, it would be expected that the undigested DNA should appear the same for each construct on the gel. F6 is also much brighter than the other plasmids, indicating a much higher concentration of DNA despite the same amount being loaded for each sample.
Figure 5-12: Analysis of the deletion constructs by restriction digest.
Lanes 1 and 11, 1 Kb Plus DNA ladder. Lane 2, undigested F1 plasmid DNA. Lane 3, SmaI-digested F1 plasmid DNA. Lane 4, SacI-digested F1 plasmid DNA. Lane 5, undigested F3 plasmid DNA. Lane 6, SmaI-digested F3 plasmid DNA. Lane 7, SacI-digested F3 plasmid DNA. Lane 8, undigested F4 plasmid DNA. Lane 9, SmaI-digested F4 plasmid DNA. Lane 10, SacI-digested F4 plasmid DNA. Lane 12, undigested F2 plasmid DNA. Lane 13, SmaI-digested F2 plasmid DNA. Lane 14, SacI-digested F2 plasmid DNA. Lane 15, undigested F5 plasmid DNA. Lane 16, SmaI-digested F5 plasmid DNA. Lane 17, SacI-digested F5 plasmid DNA. Lane 18, undigested F6 plasmid DNA. Lane 19, SmaI-digested F6 plasmid DNA. Lane 20, SacI-digested F6 plasmid DNA.

5.2.3.3 Plasmid DNA Concentration
Due to the discrepancy in the appearance and brightness of the deletion constructs, 100 ng samples of each were run alongside 100 ng samples of all of the other plasmids used in the transfection experiments for comparison. As can been seen in Figure 5-13, while the F6 plasmid DNA has a similar appearance to that of all of the control constructs, the DNA for the other five deletion constructs is much less bright, suggesting that less than 100 ng was actually loaded.
Figure 5-13: Plasmid DNA analysis

100 ng of each of the plasmids used in the transfection experiments run on a to compare the deletion construct plasmid DNA to that of the other plasmids. Lane 1, 1 Kb Plus DNA ladder. Lane 2, pGL2-basic. Lane 3, pE2F1(-242). Lane 4, pE2F1(-728). Lane 5, F1. Lane 6, F2. Lane 7, F3. Lane 8, F4. Lane 9, F5. Lane 10, F6. Lane 11, pCR3.1. Lane 12, pCMV-β-gal. Lane 13, pCMV5-Pax8. Lane 14, pCMV-E2F1.

The amounts of DNA loaded for the F1-F5 constructs were varied in order to find the amount that would show the same level of brightness on a gel as the F6 construct. Figure 5-14 shows that 2.5 times the amount of F1 and five times the amount of F2-F5 are required to achieve the same level of brightness as F6.

Figure 5-14: Deletion construct plasmid DNA analysis

Determination of the amount of each of the F1-F5 deletion constructs required to give the same brightness on an agarose gel as 100 ng of the F6 construct. Lane 1, 1 Kb Plus DNA ladder. Lane 2, 100 ng F6. Lane 3, 250 ng of F5. Lane 4, 500 ng of F5. Lane 5, 250 ng F4. Lane 6, 500 ng F4. Lane 7, 250 ng F5. Lane 8, 500 ng F5. Lane 9, 250 ng of F2. Lane 10, 500 ng of F2. Lane 11, 250 ng of F1. Lane 12, 500 ng of F1.
**Promoter-Reporter Assay**

In order to determine if this difference in concentration had an affect on the promoter-reporter assay, the deletion constructs were transfected using the modified concentrations. While all but F1 showed an increase in activity as compared to the mean of the previous assays (Figure 5-15), the activity levels were still nowhere near those of the $E2F1$ promoter constructs. Also, while the amount of deletion construct DNA was increased, the remaining plasmid amounts remained the same so that varying amounts of total DNA were transfected into the cells depending on which promoter construct was being used, and the F1-F5 plasmids were all transfected with a total DNA quantity greater than 600 ng, which was used for all of the other promoter-reporter experiments. As a test, the F1 construct was transfected with 400 ng of PAX8 both with and without 100 ng of pCR3.1 vector. The sample without the additional pCR3.1 vector had a higher activity level compared to the one that was transfected with the pCR3.1 vector (Figure 5-16). There are two possible explanations for this increase, the first being that the pCR3.1 vector may be inhibiting the luciferase assay in some way, thereby, when it is omitted from the transfection, the deletion construct is able to be fully activated. Another possibility is that when the pCR3.1 vector is included in the transfection, the total amount of DNA, at over 700 ng, is too much to be completely taken up by the cell so less deletion construct is incorporated, resulting in lower activity levels.
Figure 5-15: Deletion construct activity following overexpression of PAX8
Results of the luciferase assay using the previously determined quantities of deletion construct DNA required to obtain the same appearance on an agarose gel as the F6 construct. A) Comparison of the results of the assay using the increased quantities as compared to the mean values of the previously obtained results. B) Comparison of the results of PAX8 overexpression with the two deletion construct amounts. Error bars represent +/- 1 SEM.
Figure 5-16: F1 deletion construct activity following co-transfection with varying combinations of PAX8 and pCR3.1
Comparison of the activity levels resulting from the co-transfection of the F1 deletion construct with 0 ng of PAX8 and 500 ng of pCR3.1 vector or 400 ng of PAX8 and either 100 ng or 0 ng of pCR3.1 vector.

5.2.3.4 DNA Preparation
Since the DNA obtained from the midipreps did not appear to be functioning properly in the promoter-reporter assay, the DNA obtained from the minipreps that was used for sequencing analysis was tested. When run on a gel, the miniprep purified DNA was much brighter, with the exception of F4, than the midiprep purified DNA and all of the samples appeared as multiple bands (Figure 5-17). Since the miniprep was done without a kit, the samples were cleaned-up prior to transfection using the Qiagen PCR Clean-up Kit. Unfortunately this resulted in a complete loss of the DNA so the miniprep purified DNA was transfected as it was. Due to the problems with the F4 construct, it was omitted from this trial. As can be seen in Figure 5-18, there was a decrease in the activity of all of the promoter constructs with the co-transfection of PAX8 as compared to without. The activity levels of all of the
deletion constructs were also even lower than previously observed. Whether or not this is related to the miniprep purified DNA cannot be determined as there were also problems with the PAX8 expression construct used, as discussed in Section 5.4.

**Figure 5-17:** Deletion construct plasmid DNA purified by mini preparation. Lane 1, 1 Kb Plus DNA ladder. Lane 2, F1. Lane 3, F2. Lane 4, F3. Lane 5, F4. Lane 6, F5. Lane 7, F6.
**Figure 5-18:** Activity of miniprep purified plasmid DNA following overexpression of PAX8

Luciferase results following transfection with deletion construct plasmid DNA that was purified via miniprep, with (a) and without (b) the data from the pE2F1(-242) and (-728) plasmids.
5.3 Site-Directed Mutagenesis

In order to validate that the predicted PAX8 binding site 3 in the E2F1 promoter is the site to which PAX8 is binding and subsequently regulating the promoter activity, site-directed mutagenesis was used.

5.3.1 Cloning

5.3.1.1 Mutant Site Selection

Three sites were selected based upon the optimal PAX8 binding consensus sequence (Jun and Desplan, 1996), as shown in Figure 5-19. The two most abundant bases in the core motif (T at position 2 and C at position 3) were selected along with a third highly abundant base just downstream (G at position 6). The bases were mutated such that the most significant difference possible was made at each location. A purine was changed to a pyrimidine and vice versa, such that an A was changed to a C and a G was changed to a T (and vice versa).

![PAX8 consensus sequence and Site 3 sequence](image)

**Figure 5-19**: PAX8 consensus sequence aligned with the predicted Site 3 sequence.

Matching bases between the two sequences are shown in green. The bases making up the core motif are enclosed by the box, with those selected for mutation indicated by an asterisk.

5.3.1.2 Primers

Two pairs of primers were designed in order to incorporate the desired mutations into each strand individually and then the resulting products were combined with the two external primers in order to obtain the full-length
sequence with all three desired mutations. This is illustrated in Figure 5-20. The forward primer of set 1 and the reverse primer of set 2 were designed to incorporate the recognition sequences for the restriction enzymes SacI and HindIII, respectively, for cloning purposes.

**Figure 5-20:** Site directed mutagenesis primers
Schematic representation of the positions of the two primer sets used to incorporate the desired mutations into site 3.

### 5.3.1.3 PCR

The PCR products obtained from each of the three reactions are illustrated in Figure 5-21. Reactions 2 and 3 both required the addition of 3% DMSO whereas reaction 1 did not. The product obtained from reaction 3 using 1 µL each of 1:100 diluted products from reactions 1 and 2 was selected for cloning and the desired DNA band was purified from the agarose gel.
**Figure 5-21**: PCR products obtained using mutagenic primer pairs.

Lanes 1, 4, and 7, 1 Kb Plus DNA ladder. Lane 2, target 393 bp fragment from reaction. Lane 3, reaction 2. Lane 5, target 469 bp fragment from reaction 2 plus 3% DMSO. Lane 6, reaction 3 with pE2F1(-728) template DNA. Lane 8, reaction 3 with 1 µL each of 1/10 diluted reaction 1 and reaction 2 products plus 3% DMSO. Lane 9, reaction 3 with 1 µL each of 1/100 diluted reaction 1 and reaction 2 products plus 3% DMSO. Lane 10, reaction 3 with pE2F1(-728) template DNA plus 3% DMSO.

### 5.3.1.4 Restriction Digest

Both the mutated insert and pGL2-basic vector were digested with SacI and HindIII in preparation for ligation. The digests were incubated at 37°C for one hour, followed by a 20-minute incubation at 65°C in order to inactivate the enzymes. The digested samples were then run on a 1.5% agarose gel (Figure 5-22) and the target bands isolated.
5.3.1.5 Single Colony PCR

The insert and vector were ligated as outlined previously in section 5.2.1.5 with an incubation at room temperature for 4 hours, followed by a 10-minute incubation at 70°C to inactivate the enzyme. Following transformation, plating, and an overnight incubation at 37°C, several single colonies were selected to verify the presence and orientation of the insert using single colony PCR with the insert-specific site 3 mutagenic F1 forward primer and the vector-specific GL3R1 reverse primer. As can be seen in Figure 5-23, six of the seven colonies selected contained the insert in the proper orientation. Despite successful transformation, however, there are also faint, smaller, non-specific bands present in all of the samples in the 500 – 650 bp region. Four of the samples (Lanes 2, 3, 5, and 6) also have a smear of high molecular weight DNA traveling from the well to just beyond the insert band.
Figure 5-23: Products from single colony PCR of mutated plasmid DNA.
Lane 1, 1 Kb Plus DNA ladder. Lanes 2-7, single colonies transformed with the mutant plasmid containing the insert in the correct orientation. Lane 8, colony that was not successfully transformed. Lane 9, non-template control (NTC). The bracket indicates the size range the non-specific bands present in all of the samples.

5.3.1.6 Ligation Optimization
Due to the fact that the products from the single colony PCR were not very clean, the ligation reactions were repeated using an incubation of 24 hours at 14°C, as was used when creating the deletion constructs. The mutant plasmids were again transformed and single colony PCR was repeated, the results of which are shown in Figure 5-24. Five of the ten colonies selected contain the insert in the correct orientation.
Figure 5-24: Single colony PCR products.
Lane 1, 1 Kb Plus DNA ladder. Lanes 2-6, colonies selected from the plate of 90 µL of transformed DH5α cells. Lanes 7-11, colonies selected from the plate of 10 µL of transformed cells. Lane 12, non-template control.

5.3.1.7 Sequencing
The 500 µL culture corresponding to colony #1 (Lane 2 in Figure 5-24), was used to inoculate 50 mL of LB + Amp and grown overnight. The plasmid DNA was then purified by miniprep and sequenced in order to determine if the selected mutations were successfully incorporated. The sequence obtained was aligned with the E2F1 promoter sequence for comparison. The region in which the mutations were to be incorporated is shown in Figure 5-25, while the complete sequence alignment can be found in Appendix E. As can be seen, the desired mutations were successfully incorporated into the plasmid.

**Figure 5-25**: Mutated site 3 sequence aligned with E2F1 promoter reference sequence.
Sequencing alignment of a portion of the site 3 mutant sequence with the corresponding E2F1 promoter region using ClustalW2. Site 3 is boxed and the bases that were mutated are shown in green. Asterisks indicate matches between the two sequences.
5.3.2 Promoter-Reporter Assay

The next step was to determine if the mutated site had an effect on the ability of PAX8 to activate the $E2F1$ promoter. The mutant plasmid was used in the luciferase promoter-reporter assay, along with the pE2F1(-242) and pE2F1(-728) promoter constructs for comparison, and the pGL2-basic vector as a control. Given that the PAX8 expression construct was newly transformed and isolated for this experiment, the E2F1 expression construct was also included as a control. As can be seen in Figure 5-26, the activity of the mutant promoter was lower than that of both of the $E2F1$ promoter constructs. The results also show that the activity of all three of the promoter constructs decreased upon the expression of PAX8 and increased in the presence of E2F1.

![Figure 5-26](image)

**Figure 5-26:** Activity levels of three different $E2F1$ promoter constructs following overexpression of PAX8 and E2F1.

The two $E2F1$ promoter constructs, pE2F1(-242) and pE2F1(-728), the site 3 mutant construct, and pGL2-basic were co-transfected with either 400 ng of PAX8 or 100 ng of E2F1 expression vectors. The values shown represent the mean +/- 1 SEM.

The decreased activity level of all of the constructs in the presence of PAX8 as compared to in its absence suggests that the newly purified PAX8 expression construct is not functioning properly. Unfortunately this prevents any
conclusions being made regarding the effect of the mutated bases in site 3 on the ability for it to bind PAX8.

**5.3.3 Plasmid DNA Quality**

Given that the co-transfection of the PAX8 expression vector with the promoter constructs resulted in a decrease in activity level and the previously observed problems with the deletion construct plasmids, the next step was to again verify the quality of the plasmid DNA of both the newly purified pCMV5-Pax8 expression vector and the newly cloned site 3 mutant construct. A 100 ng amount of each was run on a 1.5% agarose gel for analysis along with some of the other plasmids used in the transfection experiments. As can be seen in Figure 5-28, both the PAX8 expression construct and the site 3 mutant appear more faintly on the gel as compared to the other plasmids, particularly the pCMV5-Pax8. These results are similar to those seen when the deletion constructs were analyzed (Figure 5-13), suggesting that the same problem has occurred once again.

![Figure 5-27: Analysis of plasmid DNA quality.](image)

100 ng of some of the plasmids used in the transfection experiments run on a 1.5% agarose gel in order to verify the quality of the DNA. Lane 1, 1 Kb Plus DNA ladder. Lane 2, pCR3.1. Lane 3, pGL2-basic. Lane 4, pE2F1(-242). Lane 5, pE2F1(-728). Lane 6, pCMV5-Pax8. Lane 7, site 3 mutant.
5.4 PAX8 Expression Construct

Since co-transfection of the newly transformed and purified pCMV5-Pax8 expression construct with the promoter constructs produced activity levels that were lower than those without an expression construct, a western blot was performed in order to verify if PAX8 was being overexpressed. As can be seen in Figure 5-29, there is a very slight increase in expression of PAX8 in the 400 ng sample as compared to the 0 ng sample. These bands were also only visible following an overnight exposure of the blot to the film. This is in contrast to the very large increase seen in the 400 ng sample versus the 0 ng sample shown in Figure 3-6 following only a 2 minute exposure of the blot. This indicates that the pCMV5-Pax8 expression vector that was transformed and purified is not expressing the PAX8 protein correctly.

![Western blot image](image)

**Figure 5-28**: Analysis of pCMV5-Pax8 expression vector.

Western blot showing the expression of PAX8 in HEK-293 cells transfected with 0 ng or 400 ng of the pCMV5-Pax8 expression vector. β-actin was used as a loading control.

5.5 Discussion

5.5.1 Deletion Constructs

The results of the single colony PCR and sequencing analysis indicate that the deletion constructs were successfully cloned. This was not, however, reflected in the results obtained from the promoter-reporter assay. The much lower
activity levels of the deletion constructs as compared to those of pE2F1(-242) and pE2F1(-728) suggest that the deletion constructs are not functioning properly. Analysis of the plasmid DNA quality, shown in Figure 5-12, reveals an obvious difference in the integrity of the DNA of the deletion constructs F1 – F5 as compared to the other plasmids, including deletion construct F6. Given that all of the samples were loaded at a quantity of 100 ng, the fact that deletion constructs F1 – F5 appear much less bright than that other plasmids suggests that a smaller quantity of DNA was actually loaded on the gel. The concentrations of all of the plasmids were determined using the same NanoDrop, suggesting that there is something in the F1 – F5 plasmid preparations that is interfering with the accurate quantification of the DNA content. If this is indeed the case, it is conceivable that whatever is inhibiting the quantification of the DNA may also be interfering with the promoter-reporter assay. The successful and accurate cloning of the deletion constructs, as evidenced by the single colony PCR and sequencing analysis results, thereby suggests that the problem arose during the plasmid preparation process. Since the midipreps were done using a kit, it is possible that one of the reagents in the kit was contaminated and subsequently contaminated the plasmid DNA during the purification process. Since all six constructs were purified using the same method, this does not explain the variation in plasmid DNA concentration and quality, particularly between the F6 construct and the other five, as seen in Figure 5-13.

Another possible contributing factor to the discrepancy between the F6 plasmid and the other five deletion constructs pertains to the very first step of the cloning procedure. The amplification of the target insert sequences from the pE2F1(-728) resulted in multiple bands for F1 through F5. The presence of non-specific PCR products necessitated that the desired band be excised from the agarose gel and then purified. The F6 PCR product, on the other hand, was a single, clean band and thus only required a clean-up of the PCR products. Given that the F6 plasmid is the only one of the six deletion constructs that has a similar appearance on an agarose gel as the other plasmids used in the transfection experiments, it leads to the question of whether that very first purification step somehow affected the activity of the other five constructs.
5.5.2 Site 3 Mutant

The same cloning method was used in the construction of the site 3 mutant as with the deletion constructs, so it is reasonable to believe that whatever is affecting the activity of the deletion constructs would most likely also affect the activity of the site 3 mutant. Whether or not the site 3 mutant is functioning properly could not be determined as it was only ever co-transfected with the new PAX8 expression vector, which did not successfully express PAX8.

5.5.3 PAX8 Expression Vector

The results of the luciferase assays in which the new PAX8 expression vector was transfected suggested that it was not functioning properly. This was confirmed by western blot, shown in Figure 5-28, where a very faint PAX8 band was only visible in the 400 ng transfected cells following an overnight exposure. This is in contrast to the results obtained with the previous expression vector that showed a large, dark band in the 400 ng sample following a two-minute exposure time. The process of obtaining more of the pCMV5-Pax8 plasmid began with the addition of 5 µL of MQ water to the tube containing the plasmid, followed by the transformation of the mixture into DH5α cells. Given the fact that no cloning was involved indicates that whatever is affecting the other plasmids most likely occurred sometime during the transformation process onward.

The possible contamination of one of the reagents from the midiprep kit seems to be the most likely source of the problem. In order to verify that this was the case, plasmid DNA would need to be purified using a different kit and then tested using the luciferase assay. Having someone else purify the plasmids using the same kit would also act to eliminate human experimental error.
CHAPTER 6
– Discussion –

6.1 Introduction

The members of the *PAX* gene family are often expressed in cancer, as extensively reviewed in the literature (Robson et al., 2006; Lang et al., 2007; Wang et al., 2008; Li and Eccles, 2012). While their expression may be aberrant or the result of a chromosomal translocation, the specific function(s) of these transcription factors in cancer remains largely unknown. It has been suggested, however, that expression of the *PAX* genes is necessary for the growth and survival of cancer cells (Muratovska et al., 2003). In support of this notion our group recently identified the cell cycle regulator gene, *E2F1*, as a potential target of *PAX8* (Li et al., 2011). The overall aim of this study was to examine this relationship in an attempt to discern whether *PAX8* directly binds to and regulates *E2F1* in cancer cells.

The *E2F1* gene was identified as being downregulated in cancer cells following siRNA treatment targeting *PAX8* (Li et al., 2011). Among the specific aims of this research were an investigation of the relationship between *PAX8* and *E2F1* in cancer cells and determination of whether *PAX8* regulates *E2F1*. This was accomplished by studying the effects of *PAX8* levels on *E2F1* promoter activity using the luciferase assay, using ChIP to determine if, and where, *PAX8* binds to *E2F1*, and generating deletion constructs to validate the functionality of each of the predicted *PAX8* binding sites. The luciferase and ChIP assays were successful and identified that *PAX8* activates the *E2F1* promoter by binding to it directly. The construction of the deletion constructs, however, was not successful and thus the functionality of each of the predicted *PAX8* binding sites could not be validated.
6.2 PAX8 Activates the E2F1 Promoter

E2F1 plays an important role in the regulation of the cell cycle at the G1/S-phase checkpoint. Bound by Rb in G0, E2F1 is then released in early G1 following phosphorylation of Rb by the CDK4/6-cyclin D complex. This free E2F1 is then able to bind to and initiate transcription of its target genes, which will drive the cell into S-phase. In addition to its role in cell cycle progression, E2F1 is also involved in mediating p53-dependent and independent apoptosis (Ginsberg, 2002; Johnson and Degregori, 2006). Following DNA damage, E2F1 protein levels are induced (Lin et al., 2001) and E2F1 localizes to the promoters of a variety of genes, including pro-apoptotic and DNA damage response genes (Muller et al., 2001; Ren et al., 2002).

Results from the luciferase assays following co-transfection of one of two E2F1 promoter constructs with a PAX8 expression vector, as presented in Chapter 3, indicated that the E2F1 promoter is responsive to PAX8 protein levels. There was only a slight difference in the relative fold change between the two promoter constructs with the pE2F1(-242) construct having a 3.7-fold (P < 0.5) increase and the pE2F1(-728) having a 4.2-fold (P > 0.5) increase following transfection with 400 ng of PAX8 expression vector as compared to 0 ng. There was, however, a large difference in amplitude with that of the pE2F1(-728) construct being ~3.5-fold higher than that of the shorter pE2F1(-242) construct. This difference in amplitude could be due to the presence of multiple binding sites, or to a nearby enhancer, but the exact cause cannot be distinguished based solely on these results.

It is possible that PAX8 does not regulate E2F1 transcription by direct binding. For example, SOCS1 (Suppressor Of Cytokine Signaling 1) was not able to activate a luciferase reporter construct under the control of the p21 promoter but instead increased the promoter activity by binding to p53, which does bind directly to the p21 promoter (Calabrese et al., 2009). While E2F1 is able to bind to the CCTα (CTP:phosphocholine cytidylyltransferase alpha) promoter driven luciferase construct, it has little effect on the activity level (Elena and Banchio, 2010). It was not until it was co-expressed with Sp1, which is also
able to directly bind the CCTα promoter and to transactivate it, that synergistic activation occurred (Elena and Banchio, 2010). Pax6 and the bone morphogenetic proteins BMP2 and BMP4 have all been shown to activate the neural cell adhesion molecule L1 driven luciferase reporter by direct binding to the HPD (homeodomain and paired domain binding site), albeit to different sequences (Meech et al., 1999). Furthermore, Pax6 is also able to bind and activate a luciferase reporter driven by its own promoter (Grocott et al., 2007). But repression of this same promoter by the TGFβ signaling pathway, on the other hand, would appear to be indirect with Smad3 physically binding to Pax6 and inhibiting its ability to bind to DNA of its own promoter, thus blocking transcription (Grocott et al., 2007).

Factors that are known to regulate the expression of E2F1 directly include Myc (Leung et al., 2008) and E2F1 (Johnson et al., 1994). PAX8 could therefore activate the E2F1 promoter through one of these factors, similar to the situation in which SOCS1 enhances transcriptional activation by binding to p53 (Calabrese et al., 2009) or as with the synergistic activation that occurs between E2F1 and Sp1 on the CCTα promoter (Elena and Banchio, 2010). Another possibility is that PAX8 could relieve negative regulation, such as the inactivation of E2F1 by the binding of Rb. PAX8 has been shown to bind to Rb (Miccadi et al., 2005) thus it is possible that PAX8 is able to bind to Rb thereby releasing E2F1 so that it can bind to and autoregulate its promoter.

It was clear that at this point in the research that there were many possibilities as to how PAX8 could regulate E2F1. Since luciferase data was not sufficient to decipher which of these mechanisms was correct, it was decided that the best way to differentiate between the various situations was to carry out assays for direct binding.
6.3 PAX8 Binds to the E2F1 Promoter

The results presented in Chapter 4 suggest that the mechanism by which PAX8 activates the E2F1 promoter is by direct binding, as indicated using ChIP. PAX8 has been shown to bind directly to several other gene promoters including the thyroid-specific genes TPO, Tg, and NIS (Zannini et al., 1992; Ohno et al., 1999). PAX8 has also been shown to directly regulate the transcription of the p53 gene, whose product, much like E2F1, is involved in cell cycle regulation and is often mutated in cancer (Stuart et al., 1995).

Zannini and colleagues (1992) identified the PAX8 binding elements in both the Tg and TPO promoters using a combination of gel shift assays, orthophenanthroline footprinting, methylation interference, and mutational analysis. The region shown to bind PAX8 in both promoters has a high degree of homology between the two sequences and both binding sites contain a conserved TGCCCAG/CT motif (Figure 6-1) (Zannini et al., 1992). The binding sites span 16 nucleotides and in both cases overlap with that of TTF-1 (Zannini et al., 1992). Despite the large amount of sequence homology between the two binding sites, PAX8 interacts with each of the promoters differently (Zannini et al., 1992). While the TGCCAG/CT motif is central to PAX8 binding, it is the surrounding nucleotides that play an important role in determining binding specificity (Zannini et al., 1992).

\[
\begin{align*}
\text{Tg} & \ -80 \ TGCCCAGTCAAGTGT \ -65 \\
\text{TPO} & \ -43 \ TGCCCATCAAGCTT \ -58
\end{align*}
\]

**Figure 6-1**: Pax8 binding site in the Tg and TPO promoters.

Comparison of the Pax8-binding sites in the Tg and TPO promoters, showing homologous bases highlighted in green and the conserved TGCCAG/CT motif in bold and underlined. Adapted from (Zannini et al., 1992).

The PAX8 binding site in the NIS promoter was identified by Ohno and colleagues (1999). They used a combination of transient transfections involving
luciferase promoter-reporter constructs, mutational analysis, DNase I footprinting and band shift assays. Two PAX8 binding sites were found in the upstream enhancer region and, as with the Tg and TPO promoters, one of the binding sites overlaps with that of TTF-1 (Ohno et al., 1999). Both sites contain the PAX8 consensus sequence (Figure 6-2) (Czerny et al., 1993; Pellizzari et al., 1996; Kozmik et al., 1997) and mutations in this region inhibit binding (Ohno et al., 1999).

\[
\begin{align*}
\text{PA} & \quad \text{CAAGCAG\_CC\_ACCG\_AG\_A\_CCTGAGTGG} \\
\text{Pax8} & \quad \text{TNNNGCGKRAVSR} \\
\text{PB} & \quad \text{GGGGTGG\_CC\_ACCG\_AG\_A\_CCTCCCTGGACATG}
\end{align*}
\]

**Figure 6-2:** Pax8 binding sites in the NIS promoter.

Alignment of the two sequences found to bind Pax8 in the upstream enhancer region of the NIS promoter (PA and PB) and the Pax8 consensus sequence (Pax8). Homologous nucleotides in the two binding sites are highlighted in green and bases that match those in the consensus sequence are underlined and in bold. Adapted from (Ohno et al., 1999).

Deletion constructs were made and mutagenesis of these was carried out, as described in Chapter 5, in an attempt to identify the motifs involved in PAX8 binding in the E2F1 promoter. Despite the fact that these experiments were unsuccessful, the ChIP assays in Chapter 4 showed that PAX8 does indeed bind to the endogenous E2F1 promoter and the results of the luciferase assays in Chapter 3 indicate that PAX8 positively regulates its activity. Therefore, while the specific binding motifs were not identified, it has been shown that PAX8 does bind to and activate the E2F1 promoter. The fact that both the pE2F1(-242) and pE2F1(-728) promoter constructs were activated by PAX8 in the luciferase assays suggests that there may be more than one motif in the E2F1 promoter region that binds PAX8, which would make the identification of these sites using deletion constructs and mutagenesis a rather complex endeavour. Gel mobility shift assays and footprinting, such as described by Zannini and
colleagues (1992) and Ohno and colleagues (1999), might have been viable alternative strategies to identify the PAX8 binding motifs in the E2F1 promoter.

Silencing of PAX8 via siRNA treatment resulted in a decrease in the activity levels of the E2F1 promoter constructs. The knockdown of PAX8 has also been shown to lead to growth retardation, the onset of senescence, and the induction of cell cycle arrest at the G1/S phase boundary in several cancer cell lines (Li et al., 2011). Together these results suggest that PAX8 plays a role in regulating the proliferation of cancer cells.

6.4 Potential PAX8 Co-Activators

In activating the transcription of certain other promoters, PAX8 interacts with co-activators. In the thyroid, for example, Pax8 interacts with Rb on the NIS promoter and with Rb and p300 on the TPO promoter (De Leo et al., 2000; Miccadei et al., 2005). In both cases, transcriptional activation is dependent on the binding of PAX8 to the promoters and not either of the co-activators. The interaction between PAX8 and its co-activators is also promoter specific, as while Rb synergistically activates transcription of both the TPO and NIS genes, it does not affect that of Tg (Miccadei et al., 2005) and acts to repress PAX8-mediated activation of E2F1 transcription (Li et al., 2011).

Both Rb and p300 are believed to act as ‘molecular matchmakers’ by assembling protein complexes (Wang, 1997). De Leo and colleagues (2000) hypothesized that Pax8 and p300 might form a complex with TBP, thereby creating a more stable complex with improved activation. This was based on the fact that p300 is known to bind TBP (Yuan et al., 1996) and that Pax5 has also been reported to interact with TBP via its partial homeodomain (Eberhard and Busslinger, 1999), a domain that is also present in PAX8. If these three proteins do indeed form a complex, it is possible that they may also act on other promoters, including E2F1.
Another transcription factor that may interact with PAX8 is Sp1. There is a highly conserved 38 bp region present in both the human and murine WT1 promoters that contains potential binding sites for Sp1, Pax2, and Pax8 and all three of these transcription factors are able to activate transcription of the gene (Hofmann et al., 1993; Dehbi and Pelletier, 1996). The sites for PAX8 and Sp1 in the WT1 promoter of both humans and mice are in very close proximity (Dehbi and Pelletier, 1996), indicating that the two transcription factors may interact. If this is the case, it is conceivable that the same situation might occur on the E2F1 promoter, as it also contains several putative Sp1 binding sites, some of which are in close proximity to the PAX8 sites.

The E2F1 gene has been identified as a target of the Myc oncoprotein (Fernandez et al., 2003). Myc dimerizes with Max and together they recognize and bind to E-box elements. A single E-box, with the sequence CACGTG has been identified in the E2F1 promoter region (Leung et al., 2008), and overlaps with the site that has been shown to bind PAX8. Given that it has been shown that Myc is often not sufficient to activate transcription on its own (Fernandez et al., 2003), and the proximity of the Myc binding site to that of PAX8, the two could form a complex in the activation of E2F1 transcription.

6.5 Future Research

6.5.1 PAX8 Binding Site Functionality

Given that analysis of the deletion constructs was not successful, an alternate method to examine the functionality of each of the predicted PAX8 binding sites would be to use site-directed mutagenesis. Unfortunately time constraints did not allow for this to be carried out. Another option would be to use gel shift and supershift assays. To rule out any discrepancies based on cloning between the deletion constructs and the 242 bp and 728 bp E2F1 promoter vectors, recloning of the two controls is suggested. The use of a different purification kit or method might also be advisable. In terms of the ChIP experiments,
optimization of the antibody amount and the specific PAX8 antibody used may aid in the identification of functional PAX8 binding sites in the E2F1 promoter.

6.5.2 Non-cancerous Cell Lines

Since this study was carried out exclusively in cancer cells, it would be interesting to examine the role of PAX8 in normal cells in order to determine if its activation of the E2F1 promoter is a normal function in healthy cells or a hijacked function in cancer cells. Measurement of endogenous E2F1 expression in any cells lines used would also be of interest.

6.5.3 PAX8 Co-activators

As discussed in Section 6.3, PAX8 often interacts with one or more co-activators when positively regulating gene transcription. Studies to identify whether PAX8 does indeed interact with other transcription factors on the E2F1 promoter, and if so, which ones, would be potentially useful.

6.5.4 Therapeutic Potential

Future research in line with this study would involve testing in vivo for potential clinical applications. If PAX8 is indeed aberrantly activating E2F1 transcription in cancer cells, leading to uncontrolled proliferation, then it would stand to reason that knockdown of PAX8 might improve prognosis. This potential is further supported by evidence that knockdown of PAX8 in several cancer cell lines results in growth retardation, triggers senescence, and induces arrest of the cell cycle (Li et al., 2011), so it would be interesting to determine if these same effects occur in vivo.

Another avenue to explore would be to find the most efficient and effective mode of delivering the siRNAs to the target cells. The use of liposomal carriers and virus-based expression vectors is limited (Libermann and Zerbini, 2006),
while lentiviral vectors have been shown to knockdown expression of target genes both \textit{in vitro} and \textit{in vivo} (Tiscornia et al., 2003).

### 6.6 Conclusion

The identification of PAX8 as an activator of \textit{E2F1} transcription in cancer cells highlights a potential role for the typically developmental transcription factor in tumourigenesis. Validation of this role \textit{in vivo} and the successful silencing of the \textit{PAX8} gene could prove to be a viable therapeutic option. Seeing as the other members of the \textit{PAX} gene family are also expressed in a variety of cancers, examining their potential involvement in cell cycle regulation would also be interesting, particularly in the case of the other two Group II family members, Pax2 and Pax5. Further studies into the role of the \textit{PAX} genes during development in normal cells as well as in cancer cells is still required in order to fully understand the roles that they play in both embryogenesis and oncogenesis.


EBIHARA, Y., MIYAMOTO, M., SHICHINOHE, T., KAWARADA, Y., CHO, Y., FUKUNAGA, A., MURAKAMI, S., UEHARA, H.,


HOL, F., GEURDS, M., CHATKUPT, S., SHUGART, Y., BALLING, R., SCHRANNDER-STUMPEL, C., JOHNSON, W., HAMEL, B. &


SANDER, M., NEUBUSER, A., KALAMARAS, J., EE, H. C., MARTIN, G. R. & GERMAN, M. S. (1997) Genetic analysis reveals that PAX6 is


VOGAN, K. J., UNDERHILL, D. A. & GROS, P. (1996) An alternative splicing event in the Pax-3 paired domain identifies the linker region as
a key determinant of paired domain DNA-binding activity. Mol Cell Biol, 16, 6677-86.


APPENDIX A
– Materials and Suppliers –

Commercial Kits
BCA Protein Assay Kit – Thermo Scientific, USA
BigDye Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems, USA
Gel Extraction Kit – Qiagen, Germany
PCR Purification Kit – Qiagen, Germany
PureLink HiPure Plasmid Filter MidiPrep Kit – Invitrogen, USA
Spin Miniprep Kit – Qiagen, Germany

Standard Reagents
1 Kb Plus DNA Ladder – Invitrogen, USA
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) – Sigma-Aldrich, USA
β-mercaptoethanol – Sigma-Aldrich, USA
Bovine Serum Albumin (BSA) – NEB, USA
Bromophenol blue – Sigma-Aldrich, USA
Complete, Mini, protease inhibitor cocktail tablets – Roche Applied Science, Germany
Dimethylsulfoxide (DMSO) – Sigma-Aldrich, USA
Ethanol absolute analytical grade – Scharlau, Spain
Ethidium bromide solution – Invitrogen, USA
Ethylene diaminetetraacetic acid (EDTA) – BDH, UK
Glacial acetic acid – BDH, UK
Glycerol – Sigma-Aldrich, USA
Hydrochloric acid (HCl) – BDH, UK
Magnesium chloride (MgCl$_2$) – BDH, UK
Methanol, analytical grade – Fisher Scientific, UK
Nonidet P-40 (NP-40) – BDH, UK
Ortho-nitrophenyl-β-galactoside (ONPG) – Sigma-Aldrich, USA
Phenylmethylsulfonylfluoride (PMSF) – Sigma-Aldrich, USA
Phosphate buffered saline (PBS) tablets – Oxoid, UK
Propan-2-ol (isopropanol) – VWR, UK
Proteinase K – Roche Applied Science, Germany
RNase A – Invitrogen, USA
Seakem LE Agarose – Cambrex, USA
Sodium acetate – BDH, UK
Sodium azide – BDH, UK
Sodium chloride (NaCl) – BDH, UK
Sodium deoxycholate – Sigma-Aldrich, USA
Sodium dodecyl sulfate (SDS) – BDH, UK
Sodium hydroxide pellets (NaOH) – BDH, UK
Sodium orthovanadate – Sigma-Aldrich, USA
Sucrose – BDH, UK
Tris(hydroxymethyl)aminomethane (Tris), Ultra Pure – Invitrogen, USA
Tween 20 – Sigma-Aldrich, USA
Xylene cyanol FF – Sigma-Aldrich, USA

**Cell Culture**

Cell Culture Lysis Reagent – Promega, USA
Dulbecco's Modified Eagle Medium (DMEM) High Glucose – Invitrogen, USA
Fetal bovine serum (FBS) – Bio International Ltd, NZ
FuGENE 6 Transfection Reagent – Roche Applied Science, Germany
Lipofectamine RNAiMAX Reagent – Invitrogen, USA
Luciferase Assay Reagent – Promega, USA
Opti-MEM I Reduced Serum Media – Invitrogen, USA
Trypsin, 0.25% with EDTA 4Na – Invitrogen, USA

**Protein Analysis**

2% Bis Solution – Bio-Rad, USA
40% Acrylamide Solution – Bio-Rad, USA
Ammonium persulfate (APS) – Sigma-Aldrich, USA
Hybond-C Extra (nitrocellulose membrane) – GE Healthcare, UK
Instant skim milk powder – Pams Foods, NZ
Kodak Medical X-ray film (MXB) General Purpose Blue – Radiographic
Supplies Ltd, NZ
MagicMark XP Standard – Invitrogen, USA
N,N,N´,N´-tetramethylethylenediamine (TEMED), Ultra Pure – Invitrogen, USA
PageRuler Prestained Protein Ladder – Fermentas, Canada
SuperSignal West Pico Chemiluminescent Substrate – Thermo Scientific, USA

**Chromatin Immunoprecipitation (ChIP)**
Chelex 100 Molecular Biology Grade Resin – Bio-Rad, USA
Dynabeads Protein G – Invitrogen Dynal AS, Norway
Formaldehyde Solution ~38% w/w – BDH, UK
Glycine – Sigma-Aldrich, USA
Lithium chloride (LiCl) – Sigma-Aldrich, USA
Platinum SYBR Green qPCR SuperMix-UDG with ROX – Invitrogen, USA
Triton-X-100 – Sigma-Aldrich, USA

**Cloning**
10 mM dNTP mix – Invitrogen, USA
Agar Bacteriological – Oxoid, UK
Ampicillin – Roche Applied Science, Germany
HindIII – New England Biolabs, USA
NEB2 Buffer - New England BioLabs, USA
NheI – New England BioLabs, USA
Phusion High-Fidelity DNA polymerase – Finnzymes, Finland
Phusion Hot Start II DNA Polymerase – Finnzymes, Finland
SacI – Roche Applied Science, Germany
SmaI – Boehringer Manheim, Germany
T4 DNA ligase – Gibco Brl Life Technologies
T4 DNA ligase buffer – Gibco Brl Life Technologies
Tryptone – Oxoid, UK
Yeast extract – Oxoid, UK
## APPENDIX B

### – Buffers and Solutions –

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0.5 M Tris-HCl solution, pH 6.8</strong></td>
<td>0.5 M Tris in dH$_2$O, Adjust pH to 6.8 with concentrated HCl</td>
</tr>
<tr>
<td><strong>1.5 M Tris-HCl solution, pH 8.8</strong></td>
<td>1.5 M Tris in dH$_2$O, Adjust pH to 8.8 with concentrated HCl</td>
</tr>
<tr>
<td><strong>1.5% agarose gel</strong></td>
<td>1.5 g agarose, 1x TAE + EtBr to 100 mL</td>
</tr>
<tr>
<td><strong>10% Chelex-100</strong></td>
<td>10% Chelex-100 (w/v) in sterile water (MQ or distilled)</td>
</tr>
<tr>
<td><strong>10% separating gel</strong></td>
<td>10% - 40% acrylamide, 0.33% - 2% bis acrylamide solution, 25% - 0.4% SDS + 1.5 M Tris-HCl, pH 8.8, 0.1% - 10% (w/v) APS in dH$_2$O, 0.05% TEMED in dH$_2$O</td>
</tr>
<tr>
<td><strong>10x TAE buffer</strong></td>
<td>40 mM Tris, 20 mM glacial acetic acid, 0.025 mM EDTA in dH$_2$O</td>
</tr>
<tr>
<td><strong>1x SDS running buffer</strong></td>
<td>25 mM Tris, 192 mM glycine, 0.1% SDS in dH$_2$O</td>
</tr>
<tr>
<td><strong>1x TAE buffer + EtBr</strong></td>
<td>200 mL 10x TAE buffer, dH$_2$O to 2 L, 50 µL ethidium bromide</td>
</tr>
<tr>
<td><strong>1x western transfer buffer</strong></td>
<td><strong>4x protein sample buffer</strong></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>20% methanol</td>
<td>0.24 M Tris-HCl, pH 6.8</td>
</tr>
<tr>
<td>0.01% SDS</td>
<td>0.24 M SDS</td>
</tr>
<tr>
<td>in 1x SDS running buffer</td>
<td>40% glycerol</td>
</tr>
<tr>
<td>Prepare fresh and store at -20°C until use.</td>
<td>20% β-mercaptoethanol</td>
</tr>
<tr>
<td></td>
<td>0.02% bromophenol blue</td>
</tr>
<tr>
<td></td>
<td>in dH₂O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>5% stacking gel</strong></th>
<th><strong>6x DNA loading dye</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>5% - 40% acrylamide</td>
<td>0.25% bromophenol blue</td>
</tr>
<tr>
<td>0.17% - 2% bis acrylamide solution</td>
<td>0.25% xylene cyanol FF</td>
</tr>
<tr>
<td>25% - 0.4% SDS + 0.5 M Tris-HCl, pH 6.8</td>
<td>40% sucrose</td>
</tr>
<tr>
<td>0.1% - 0% (w/v) APS in dH₂O</td>
<td>in MQ-H₂O</td>
</tr>
<tr>
<td>0.125% TEMED</td>
<td></td>
</tr>
<tr>
<td>in dH₂O</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>70% Ethanol</strong></th>
<th><strong>95% Ethanol</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>70% absolute ethanol</td>
<td>95% absolute ethanol</td>
</tr>
<tr>
<td>in MQ-H₂O</td>
<td>in MQ-H₂O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Alkaline SDS</strong></th>
<th><strong>β-galactosidase substrate</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2 M NaOH</td>
<td>2 mM MgCl₂</td>
</tr>
<tr>
<td>10% - 10% (w/v) SDS in dH₂O</td>
<td>100 mM β-mercaptoethanol</td>
</tr>
<tr>
<td>in dH₂O</td>
<td>1.33 mg/mL ONPG</td>
</tr>
<tr>
<td>Made fresh prior to each use.</td>
<td>in PBS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Cell lysis buffer</strong></th>
<th><strong>ChIP dilution buffer</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1x complete mini protease inhibitor cocktail</td>
<td>2 mM EDTA</td>
</tr>
<tr>
<td>1 mM PMSF</td>
<td>100 mM NaCl</td>
</tr>
<tr>
<td>1 mM sodium orthovanadate in RIPA buffer</td>
<td>20 mM Tris-HCl, pH 8.0</td>
</tr>
<tr>
<td></td>
<td>0.5% Triton-X-100</td>
</tr>
<tr>
<td></td>
<td>in dH₂O</td>
</tr>
</tbody>
</table>
### ChIP wash buffer I (low salt)
- 2 mM EDTA
- 20 mM Tris-HCl, pH 8.0
- 0.1% SDS
- 1% Triton-X-100
- 150 mM NaCl
- in dH₂O

### ChIP wash buffer II (high salt)
- 2 mM EDTA
- 20 mM Tris-HCl, pH 8.0
- 0.1% SDS
- 1% Triton-X-100
- 500 mM NaCl
- in dH₂O

### ChIP wash buffer III (LiCl)
- 1 mM EDTA
- 10 mM Tris-HCl, pH 8.0
- 1% NP-40
- 1% sodium deoxycholate
- 0.25 M LiCl
- in dH₂O

### LB agar + Amp
- 1.5% agar
- in LB broth
- Sterilize by autoclaving
- Allow to cool and add ampicillin (100 µg/mL)

### LB broth + Amp
- 1% tryptone
- 0.5% yeast extract
- 0.5% sodium chloride
- in dH₂O
- Sterilize by autoclaving.
- Allow to cool and add ampicillin (100 µg/mL)

### Lysis buffer (miniprep)
- 0.025 M Tris, pH 8
- 0.01 M EDTA, pH 8
- 0.05 M sucrose (stored at 4°C)
- Add 0.1 mg/mL RNAse A prior to use.
<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP-40 lysis buffer</td>
<td>20 mM HEPES, pH 7.5</td>
</tr>
<tr>
<td></td>
<td>150 mM NaCl</td>
</tr>
<tr>
<td></td>
<td>10% glycerol</td>
</tr>
<tr>
<td></td>
<td>1% NP-40</td>
</tr>
<tr>
<td></td>
<td>2 mM EDTA</td>
</tr>
<tr>
<td></td>
<td>Add the following protease inhibitors prior to use:</td>
</tr>
<tr>
<td></td>
<td>1x complete mini protease inhibitor cocktail</td>
</tr>
<tr>
<td></td>
<td>1 mM PMSF</td>
</tr>
<tr>
<td></td>
<td>1 mM sodium orthovanadate</td>
</tr>
<tr>
<td>PBS</td>
<td>One PBS tablet</td>
</tr>
<tr>
<td></td>
<td>dH$_2$O to 100 mL</td>
</tr>
<tr>
<td>PBST</td>
<td>0.1% Tween-20</td>
</tr>
<tr>
<td></td>
<td>in PBS</td>
</tr>
<tr>
<td>RIPA buffer</td>
<td>0.05 M Tris-HCl, pH 8</td>
</tr>
<tr>
<td></td>
<td>0.15 M NaCl</td>
</tr>
<tr>
<td></td>
<td>1% NP-40</td>
</tr>
<tr>
<td></td>
<td>0.5% sodium deoxycholate</td>
</tr>
<tr>
<td></td>
<td>0.1% SDS</td>
</tr>
<tr>
<td></td>
<td>MQ-H$_2$O, and stored at 4°C</td>
</tr>
<tr>
<td>SDS lysis buffer</td>
<td>50 mM - 1 M Tris-HCl, pH 8.1</td>
</tr>
<tr>
<td></td>
<td>1% - 10% (w/v) SDS in dH$_2$O</td>
</tr>
<tr>
<td></td>
<td>10 mM - 0.5 M EDTA in dH$_2$O</td>
</tr>
<tr>
<td></td>
<td>Add the following protease inhibitors prior to use:</td>
</tr>
<tr>
<td></td>
<td>1x complete mini protease inhibitor cocktail</td>
</tr>
<tr>
<td></td>
<td>1 mM PMSF</td>
</tr>
<tr>
<td></td>
<td>1 mM sodium orthovanadate</td>
</tr>
<tr>
<td>TE</td>
<td>1 mM EDTA</td>
</tr>
<tr>
<td></td>
<td>10 mM - 1 M Tris-HCl, pH 8.0</td>
</tr>
<tr>
<td></td>
<td>in dH$_2$O</td>
</tr>
<tr>
<td>Western blocking buffer</td>
<td>2.5% instant skim milk powder</td>
</tr>
<tr>
<td></td>
<td>in PBST</td>
</tr>
</tbody>
</table>
APPENDIX C
– ChIP-qPCR Sequencing Chromatograms –

Site 1 – Forward primer
Sequencing Reaction #1

Site 1 – Reverse primer
Sequencing Reaction #1
Site 3 – Reverse primer

Sequencing Reaction #1

Sequencing Reaction #2

Sequencing Reaction #3
Sequencing Reaction #4

Site 4 – Forward primer

Sequencing Reaction #1
Site 4 – Reverse primer

Sequencing Reaction #1

Sequencing Reaction #2

Sequencing Reaction #3
Site 5 – Forward primer
Sequencing Reaction #1

Site 5 – Reverse primer
Sequencing Reaction #1
Site 6 – Forward primer

Sequencing Reaction #1

Sequencing Reaction #2
Site 6 – Reverse primer

Sequencing Reaction #1

Sequencing Reaction #2
Sequencing Reaction #3

Sequencing Reaction #4
Control Site – Forward primer
Sequencing Reaction #1

Control Site – Reverse primer
Sequencing Reaction #1
ClustalW2 was used to align the reference sequence for each deletion construct with the sequencing results obtained using the GL3F2 and GL3R1 primers (in bold) that bind to the pGL2-basic vector outside of the multiple cloning site. The NheI (GCTAGC) and HindIII (AAGCTT) sequences used to clone the inserts into the vector are highlighted in yellow and the predicted PAX8 binding sites are highlighted in green. The SmaI (CCCGGG) and SacI (GAGCTC) sites that were used to check for the presence of deletion mutants are highlighted in blue. The transcription start site is indicated by +1 and the bases are numbered relative to that position. Asterisks indicate matches and dashes indicate gaps in the sequence.

F1 ref. seq.  CTGCATTTCTAGTTGTGGTCTTCCTGCAACTCTCATCAATGTATCTTATGGTACTGTAACTGAG -645
F1 -------------------------------------------CTYCCYYMTGT-TCTTATGGTACTGTAACTGAG
** **  **************************************

F1 ref. seq.  CTAACATTAACCCGGAGGTACCTTAGGCTGTGCCTCTTCTTCTCGCTCTACG
F1 CTAACATTAACCCGGAGGTACCTTAGGCTGTGCCTCTTCTTCTCGCTCTACG
** ******************************* **  *****

F1 ref. seq.  CCCGAATCCCCCTCGACATTTCCAGGCACCCTCTCTTCAGCTTGCCCT
F1 CCCGAATCCCCCTCGACATTTCCAGGCACCCTCTCTTCAGCTTGCCCT
************************************************************

F1 ref. seq.  GATAGGGCTGGGTGCAGGATTAGGATAATGTCATGGGTGAGGCAAGTTGAGGATGGAAGA
F1 GATAGGGCTGGG- GCAAGTTGAGGATGGAAGA
************  *********************************** ***********

F1 ref. seq.  GGTGGCTGATGGCTGGGCTGTGGAACTGATGATCCTGAAAAGAAGGGGACAGTCTCTGT
F1 GGTGGCTGATGGCTGGGCTGTGGAACTGATGATCCTGAAAAGAAGGGGACAGTCTCTGT
***************************************************** ****

F1 ref. seq.  GAAATCTAAGCTGAGGCTGTTGGGGCCTACAGGTTGAGGGTCACGTGCAGAAGAGAGGCT
F1 GAAATCTAAGCTGAGGCTGTTGGGGCCTACAGGTTGAGGGTCACGTGCAGAAGAGAGGCT
************************************************************

F1 ref. seq.  GGCCTATGTTCCGGGTCCCCACGCCTCCAGGGAAGAGGGGACGCCCGGGCTGGGGCGGG
F1 GGCCTATGTTCCGGGTCCCCACGCCTCCAGGGAAGAGGGGACGCCCGGGCTGGGGCGGG
************************************************************
179
APPENDIX E
– Site 3 Mutant Sequence Alignment –

ClustalW2 was used to align the E2F1 reference sequence with the sequencing results obtained using the Site 3 Mutagenic F1 and R2 primers (in bold). The SacI (GAGCTC) and HindIII (AAGCTT) sequences used to clone the insert into the vector are highlighted in blue. The predicted PAX8 binding sites are highlighted in green, with the mutations inserted into site 3 highlighted in yellow. The transcription start site is indicated by +1 and the bases are numbered relative to that position. Asterisks indicate matches and dashes indicate gaps in the sequence.
E2F1 ref. seq.  TCACGGCCGGGCGGCCAATTGTCGGGCTCGGCGGACTTTGCAGGCAGCGGCGGCCGGGGGCGGAGC
M3  TCACGCACGCGGCCCAATTGTCGGGCTCGGCGGACTTTGCAGGCAGCGGCGGCCGGGGGCGGAGC

***** ***************************************

+1

E2F1 ref. seq.  AAGGATTTGGCGCGTGAAAAGTGGCCGGGACTTTGCAGGCAGCGGCGGCCGGGGGCGGAGC
M3  AAGGATTTGGCGCGTGAAAAGTGGCCGGGACTTTGCAGGCAGCGGCGGCCGGGGGCGGAGC

****** ***************************************

E2F1 ref. seq.  GGGATCGAGCCCTCGGCGGCTGGCTCATCGC
M3  GGGATCGAGCCCTCGGCGGCTGGCTCATCGC

*****************************

E2F1 ref. seq.  GGGATCGAGCCCTCGGCGGCTGGCTCATCGC
M3  GGGATCGAGCCCTCGGCGGCTGGCTCATCGC

*****************************

E2F1 ref. seq.  GGGATCGAGCCCTCGGCGGCTGGCTCATCGC
M3  GGGATCGAGCCCTCGGCGGCTGGCTCATCGC

*****************************