Evolutionary Comparison In The Vertebrate Lineage of \textit{WT1}, a Wilms' Tumour Predisposition Gene

Gillian R.L. Kent

Ph.D.
University of Edinburgh
and
MRC Human Genetics Unit

1994
Declaration

I declare

a) This thesis was composed by myself alone

b) The work herein is my own unless clearly stated
In memory of my grandmother

Lavinia Paterson
Acknowledgements

There are many people without whom this project would never have got off the ground. Working on so many species would not have been possible without very many generous gifts from around the world. Hopefully I have managed to mention all of these within the text but there are several people I would particularly like to thank.

Anne Marie Coriat for the alligator in situ data.
Trevor Jowett for the zebrafish material and also allowing me to spend a week in a contamination-free lab.
Stephen Wilcox and Jenny Marshall Graves for screening the marsupial library.
David Housman, Daniel Haber and Alan Buckler for the mouse and human WT1 cDNA clones.
P.A. Tsonis for sending me the axolotl WT1 sequence.

During the last three and a half years at the MRC there have been an immense number of kind people who have been very generous with both their time and their advice. Among those I probably pestered the most, pinched their solutions, borrowed their books etc. are;

Duncan Davidson, Jonathan Bard, Bob Hill for trying to teach me development.
Isabel Hanson, Ian Jackson, Wendy Bickmore, Anne Seawright, Andreas Schedl, Cathy Abbott for excellent technical advice.
Lesley McInnes for teaching me the whole-mount technique.
John Inglis for advice on degenerate PCR and evolution analysis.

Also without the help of the following I probably would have lost a lot more sleep over this thesis.

Kathy Williamson and Duncan for initiating me into the mysteries of Papayrus
Andreas, Sarah Danes, James Williams for proofreading.
Douglas, Norman and Sandy in Photography.
Justin for always finding one more interesting article.
Also the whole lab. for putting up with me while I've been writing.

Finally I've been lucky enough to have some excellent supervisors.

Andy Leigh-Brown for being my University supervisor.
Veronica van Heyningen for encouragement, ideas, keeping my English at bay and generally trying to mother us all.
Nick Hastie for always wanting more data!
Abstract

The \textit{WTI} gene was isolated by positional cloning as a candidate gene implicated in predisposition to the paediatric kidney cancer Wilms' tumour. Because of its early onset and histological resemblance to immature kidneys, Wilms' tumour was thought to arise from an aberration in the normal developmental pathway. From the amino acid sequence, \textit{WTI} was predicted to be a transcription factor with an N-terminal proline/glutamine rich transregulatory domain and four C-terminal \textit{TFIIIA}-like zinc fingers. Two alternative splices have been found in the transcript, resulting in the insertion of 17aa or 3aa (KTS). Functional analysis has shown that \textit{WTI} is capable of sequence specific DNA binding and of regulating transcription. Analysis of \textit{WTI} mRNA in the developing embryo has shown a pattern of expression consistent with an important role in nephrogenesis, specifically an involvement in mesenchymal-epithelial cell transition. \textit{WTI} analysis in Wilms' tumours revealed that both functional copies are lost in about 10\% of sporadic Wilms' tumours. This confirmed the role of \textit{WTI} as a tumour suppressor gene in at least some cases, (following Knudson's two hit hypothesis for tumorigenesis). Predisposition to Wilms' tumour can also be associated with developmental abnormalities, including genitourinary malformations. Expression of \textit{WTI} is seen in the earliest stages of gonad development and constitutionally heterozygous mutations in \textit{WTI} are frequently associated with genitourinary abnormalities. The most severe abnormalities, in Denys-Drash syndrome, have been linked to heterozygous constitutional \textit{WTI} missense mutations in the zinc fingers. The cloning of the mouse \textit{Wt1} gene revealed a very high level of similarity to the human gene, as well as a very similar pattern of expression during development. The major problem with the use of this system as a model is that Wilms' tumours have never been observed in mouse, even in the Sey(Dey) mouse which is hemizygous for \textit{Wt1}.

In order to complement the current lines of investigation it was decided to attempt to isolate a number of \textit{WTI} orthologs from a range of vertebrates to produce wider sequence comparison and eventually a functional comparison. Evolutionary analysis is a particularly appropriate line of investigation for kidney development because across all the vertebrates nephrogenesis seems to occur in a similar manner and kidney tumours which histologically resemble Wilms' tumours have been found in many vertebrates. Also mammalian kidney development passes through three stages. In the lower vertebrates only the first two stages occur so that the human foetal kidney resembles the adult kidney of the anamniotes. In all the vertebrates the close association of the urinary and gonadal systems is maintained, with variations.
Therefore the investigation of WTI expression in these different but similar systems may point to important roles for WTI. It will also produce more manipulable animal models in which the roles of many other developmental genes are now being investigated.

The conservation of sequences homologous to WTI in a range vertebrates was first demonstrated by zoo blot hybridisation with human and mouse WTI probes. Partial clones of WTI orthologs from rat, pig, chick, alligator, *Xenopus laevis* and zebrafish were then isolated by cross-species PCR, using degenerate primers designed on the basis of the mouse and human sequences. These were then used as probes to isolate larger regions of the gene from cDNA libraries. Clones for marsupial mouse, chick, alligator and zebrafish were obtained. Comparison of these has shown that there is a high level conservation of the predicted protein sequence due to selection pressure. The highest level of conservation is found in the DNA binding zinc fingers, which is consistent with that observed in comparisons of other transcription factors. Investigation of the conservation of the two alternative splice forms of the gene indicated that the KTS alternative splice which is involved in the modulation of DNA binding is very probably conserved in zebrafish, but the 17 amino acid alternative splice was not observed outside the eutherian mammals and was also adjacent to one of the most highly substituted regions of the protein indicating that this could be involved in class specific development. Much of the transregulatory domain was highly conserved, the exceptions being two homopolymer motifs indicating that these may have arisen due to instability of a trinucleotide repeat and may not be functionally important. The mutations so far described in WTI have all, bar one, been located at conserved positions.

The expression of WTI in chick, alligator and *Xenopus* has been investigated by Northern blotting and whole mount in situ hybridisation and the expected pattern during kidney development seen, consistent with a conserved function. Expression was also observed in lateral plate mesoderm and ectodermal derived structures which has also been observed in mammals.

Therefore WTI is a highly conserved gene at both the structural and expression level, indicating an evolutionary conserved role in development. This study paves the way for a detailed comparison of genitourinary development and Wilms' like tumours in the vertebrates.
# Table of Contents

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1 Introduction</td>
<td></td>
</tr>
<tr>
<td>1.1. Introduction</td>
<td>20</td>
</tr>
<tr>
<td>1.2. Wilms' tumour and the identification of a putative predisposition gene-WT1</td>
<td>21</td>
</tr>
<tr>
<td>1.2.1 Incidence of Wilms' tumour</td>
<td>21</td>
</tr>
<tr>
<td>1.2.2 Knudson and Strong two hit model for tumorigenesis.</td>
<td>23</td>
</tr>
<tr>
<td>1.2.3 Mapping of the WAGR locus to 11p13</td>
<td>24</td>
</tr>
<tr>
<td>1.2.4 Isolation of WT1</td>
<td>24</td>
</tr>
<tr>
<td>1.3 Structure and function of WT1</td>
<td>26</td>
</tr>
<tr>
<td>1.3.1 Zinc fingers</td>
<td>26</td>
</tr>
<tr>
<td>1.3.2 Proline/glutamine rich region</td>
<td>26</td>
</tr>
<tr>
<td>1.3.3 Alternative splices</td>
<td>29</td>
</tr>
<tr>
<td>1.3.4 Identification of the WT1 protein</td>
<td>30</td>
</tr>
<tr>
<td>1.3.5 Transcription factor activity</td>
<td>30</td>
</tr>
<tr>
<td>1.3.5.1 DNA binding</td>
<td>30</td>
</tr>
<tr>
<td>1.3.5.2 Target promoters</td>
<td>30</td>
</tr>
<tr>
<td>1.3.5.3 Mapping of functional domains</td>
<td>31</td>
</tr>
<tr>
<td>1.3.5.4 Protein-protein interactions and modulation of WT1 activity</td>
<td>32</td>
</tr>
<tr>
<td>1.3.5.5 Regulation of WT1 expression</td>
<td>32</td>
</tr>
<tr>
<td>1.4 Evidence that WT1 is a predisposition gene for Wilms' tumour</td>
<td>33</td>
</tr>
<tr>
<td>1.4.1 Mapping deletions in Wilms' tumours to the WT1 locus</td>
<td>33</td>
</tr>
<tr>
<td>1.4.2 Intragenic deletions</td>
<td>33</td>
</tr>
<tr>
<td>1.5 Kidney development and WT1</td>
<td>36</td>
</tr>
<tr>
<td>1.5.1 The developing kidney</td>
<td>36</td>
</tr>
<tr>
<td>1.5.2 The molecular basis of kidney differentiation</td>
<td>40</td>
</tr>
<tr>
<td>1.5.3. Wilms' tumours</td>
<td>43</td>
</tr>
<tr>
<td>1.6 WT1 and intermediate mesoderm development</td>
<td>44</td>
</tr>
<tr>
<td>1.7 WT1 expression outside the intermediate mesoderm</td>
<td>47</td>
</tr>
<tr>
<td>1.8 Involvement of WT1 in genitourinary abnormalities</td>
<td>48</td>
</tr>
<tr>
<td>1.8.1 Mapping of genitourinary abnormalities to the WT1 region</td>
<td>48</td>
</tr>
<tr>
<td>1.8.2 Denys-Drash Syndrome</td>
<td>49</td>
</tr>
<tr>
<td>1.8.3 Dominant-negative hypothesis</td>
<td>49</td>
</tr>
<tr>
<td>1.9 Involvement in other developmental abnormalities</td>
<td>52</td>
</tr>
<tr>
<td>1.10 The WT1 knockout mouse</td>
<td>53</td>
</tr>
<tr>
<td>1.11 Is WT1 the only gene responsible for predisposition to Wilms' tumour?</td>
<td>53</td>
</tr>
<tr>
<td>1.11.1 Lack of mutations in WT1</td>
<td>53</td>
</tr>
<tr>
<td>1.11.2 Dominant mutations and trans-splicing effects</td>
<td>54</td>
</tr>
<tr>
<td>1.11.3 A possible role for imprinting</td>
<td>54</td>
</tr>
<tr>
<td>1.11.4 A second Wilms' tumour locus on 11p</td>
<td>55</td>
</tr>
<tr>
<td>1.11.5 Not all Wilms' tumour predisposition loci map to chromosome 11</td>
<td>56</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1.11.6 Progression of Wilms' tumours</td>
<td>56</td>
</tr>
<tr>
<td>1.12 <strong>Evolutionary comparison of the vertebrate genitourinary system</strong></td>
<td>56</td>
</tr>
<tr>
<td>1.12.1 The vertebrate kidney</td>
<td>57</td>
</tr>
<tr>
<td>1.12.2 Amniote kidney development</td>
<td>57</td>
</tr>
<tr>
<td>1.12.3 Anamniote kidney development</td>
<td>60</td>
</tr>
<tr>
<td>1.12.4 The metanephros</td>
<td>60</td>
</tr>
<tr>
<td>1.12.5 The gonads and genital ducts</td>
<td>60</td>
</tr>
<tr>
<td>1.13 <strong>Evolutionary comparison of Wilms'-like tumours</strong></td>
<td>62</td>
</tr>
<tr>
<td>1.13.1 Mammals</td>
<td>62</td>
</tr>
<tr>
<td>1.13.2 Birds</td>
<td>63</td>
</tr>
<tr>
<td>1.13.3 Anamniotes</td>
<td>64</td>
</tr>
<tr>
<td>1.14 <strong>Evolutionary comparison as a method for defining gene function</strong></td>
<td>64</td>
</tr>
<tr>
<td>1.14.1 Structural analysis</td>
<td>64</td>
</tr>
<tr>
<td>1.14.2 Functional analysis</td>
<td>66</td>
</tr>
<tr>
<td>1.14.3 Animal models</td>
<td>67</td>
</tr>
<tr>
<td>1.15 Conclusion and aims of the project</td>
<td>67</td>
</tr>
</tbody>
</table>

**Chapter 2 Materials and methods**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 <strong>Bacterial manipulation</strong></td>
<td>70</td>
</tr>
<tr>
<td>2.1.1 Genotypes</td>
<td>70</td>
</tr>
<tr>
<td>2.1.1.1 Cell strains</td>
<td>70</td>
</tr>
<tr>
<td>2.1.1.2 Bacteriophage</td>
<td>70</td>
</tr>
<tr>
<td>2.1.2 Electro-transformation of <em>E.coli</em></td>
<td>70</td>
</tr>
<tr>
<td>2.1.2.1 Preparation of electro-competent cells</td>
<td>70</td>
</tr>
<tr>
<td>2.1.2.2 Electro-transformation</td>
<td>71</td>
</tr>
<tr>
<td>2.1.3 Selection of cells harbouring recombinant plasmids</td>
<td>71</td>
</tr>
<tr>
<td>2.1.4 Bacterial colony lifts</td>
<td>71</td>
</tr>
<tr>
<td>2.1.5 Small scale preparation of plasmid DNA</td>
<td>72</td>
</tr>
<tr>
<td>2.1.5.1 Method 1</td>
<td>72</td>
</tr>
<tr>
<td>2.1.5.2 Method 2</td>
<td>72</td>
</tr>
<tr>
<td>2.1.6 Large scale preparation of plasmid DNA</td>
<td>73</td>
</tr>
<tr>
<td>2.1.7 Glycerol stocks of bacteria</td>
<td>74</td>
</tr>
<tr>
<td>2.1.8 Isolation of clones from bacteriophage libraries</td>
<td>74</td>
</tr>
<tr>
<td>2.1.8.1 Preparation of host cells for bacteriophage</td>
<td>74</td>
</tr>
<tr>
<td>2.1.8.2 Titering of libraries</td>
<td>74</td>
</tr>
<tr>
<td>2.1.8.3 Screening libraries</td>
<td>75</td>
</tr>
<tr>
<td>2.1.8.4 Amplification of phage by the plate lysate method</td>
<td>76</td>
</tr>
<tr>
<td>2.1.8.5 <em>In vivo</em> excision of λZAP phagemids</td>
<td>76</td>
</tr>
<tr>
<td>2.1.8.6 Purification of λgt11 DNA using LambdaSorb™</td>
<td>77</td>
</tr>
<tr>
<td>2.2 <strong>DNA/RNA manipulations</strong></td>
<td>77</td>
</tr>
<tr>
<td>2.2.1 DNA preparation from tissue</td>
<td>77</td>
</tr>
<tr>
<td>2.2.2 RNA preparation from tissue or cell lines</td>
<td>78</td>
</tr>
<tr>
<td>2.2.3 Agarose gel electrophoresis</td>
<td>78</td>
</tr>
<tr>
<td>2.2.3.1 DNA electrophoresis</td>
<td>78</td>
</tr>
<tr>
<td>2.2.3.2 RNA electrophoresis</td>
<td>79</td>
</tr>
</tbody>
</table>
2.2.4 Vacuum transfer of DNA and RNA 79
2.2.5 Hybridisation 80
2.2.6 Purification of DNA 81
2.2.6.1 Phenol:chloroform extraction 81
2.2.6.2 GENECELEAN® II (Bio 101) 81
2.2.6.3 Purification of a specific DNA fragment 81
2.2.6.3.1 GENECELEAN® II (Bio 101) 81
2.2.6.3.2 Spin-X 82
2.2.6.4 "Double GENECELEAN®" protocol for cloning PCR products 82
2.2.7 Quantification of DNA fragments 82
2.2.8 Enzyme modification of DNA and RNA 82
2.2.8.1 Restriction digest of plasmid DNA 82
2.2.8.2 Restriction digests of genomic DNA 83
2.2.8.3 Production of radiolabelled DNA probes 83
2.2.8.3.1 Random primed labelling 83
2.2.8.3.2 Specific primed labelling 84
2.2.8.4 Preparation of riboprobes 84
2.2.8.5 cDNA synthesis 85
2.2.8.6 Modifying ragged ends of PCR products 86
2.2.8.7 Ligation of DNA fragments into vectors 86
2.2.8.7.1 Cohesive end subcloning 86
2.2.8.7.2 Blunt ended ligation of PCR products 87
2.2.8.7.3 Turbo cloning of PCR products 87
2.3 Sequencing DNA 87
2.3.1 Double stranded sequencing of plasmid DNA 88
2.3.2 Direct sequencing of PCR products 88
2.3.3 Polyacrylamide gel electrophoresis (PAGE) of sequencing reactions 89
2.4 Detection of radioactive signal 90
2.4.1 Autoradiography 90
2.4.2 PhosphorImager analysis 90
2.5 Incubation and analysis of embryos 90
2.5.1 Rearing of Xenopus embryos 90
2.5.2 Whole mount in situ hybridisation 91
2.5.2.1 Incubation and dissection of embryos 91
2.5.2.2 Preparation of embryos 91
2.5.2.3 Hybridisation 92
2.5.2.4 Antibody conjugate binding and staining 92
2.5.2.5 Photography of embryos 93
2.6 Polymerase chain reaction (PCR) 93
2.6.1 Synthesis and purification of oligonucleotides 93
2.6.2 Reaction conditions 93
2.6.3 PCR template 94
2.7 Solutions 94
2.8 Computer analysis of DNA sequences 98
Chapter 3 Identification of sequences cross hybridising to WTI in the vertebrates

3.1 Introduction 100
3.2 Identification of sequences cross hybridising to WTI in mammals 100
3.3 Detection of sequences cross hybridising to WTI in non mammalian vertebrates 102
3.4 Conclusions 102

Chapter 4 Isolation of WTI orthologs

4.1 Isolation by cross species PCR 106
4.1.1 Introduction 106
4.1.2 Experimental design 108
4.1.2.1 Template 108
4.1.2.2 Primer design 108
4.1.2.3 Reaction conditions 112
4.1.3 Testing PCR primers 112
4.1.4 Isolation of rat, pig and alligator WTI orthologs 112
4.1.4.1 Cross species PCR 112
4.1.4.2 Cloning PCR products 114
4.1.5 Redesign of PCR primers 121
4.1.6 Isolation of chick, Xenopus and axolotl WTI orthologs 121
4.1.7 Isolation of the zebrafish WTI ortholog 121
4.1.7.1 Degenerate PCR 121
4.1.7.2 Turbo cloning 123
4.2 Isolation by library screening 126
4.2.1 Introduction 126
4.2.2 Library screens 128
4.2.3 WTI cDNA clones isolated 128
4.2.4 Analysis of zinc finger 4 in chick 132
4.2.5 Isolation of zebrafish WTI genomic clone (Z1a) 136
4.3 Discussion 136

Chapter 5 Structural and evolutionary comparison of WTI orthologs

5.1 Introduction 141
5.2 Structural analysis 141
5.2.1 Sequence motifs 141
5.2.1.1 Structure prediction 141
5.2.1.2 Post-translational modification 144
5.2.1.3 Exon structure 144
5.2.2 5' Non coding region 144
5.2.3 The predicted initiation codon is conserved 146
5.2.4 Transregulatory domain 146
5.2.4.1 Homopolymer domains
5.2.4.2 The amino acid composition is similar to other transregulatory domains
5.2.4.3 Mutations
5.2.5 The conservation of the putative leucine zipper in exon 4
5.2.6 Exon 5 - the 17aa alternative splice is not conserved in the reptilian lineage
5.2.7 The divergence and conservation of exon 6
5.2.8 The zinc finger region is the most highly conserved region of WTI
5.2.9 The KTS alternative splice has been conserved through vertebrate evolution
5.2.10 The C-terminal region is not changed in mouse
5.2.11 3' Non coding region
5.3 Evolutionary analysis
5.3.1 Rate of sequence evolution of WTI
5.3.1.1 Introduction
5.3.1.2 Analysis of WTI
5.3.2 Phylogenetic analysis

Chapter 6 Investigation of the expression of WTI

6.1 Introduction
6.2 Northern analysis of WTI mRNA
6.3 In situ hybridisation
6.3.1 Comparison of the developmental stages of amniotes
6.3.2 WTI expression in chick embryos
6.3.3 WTI expression in alligator embryos
6.3.4 Xenopus whole-mount hybridisation
6.4 Comparison of WTI expression

Chapter 7 Discussion

7.1 Evolution of transcription factors in vertebrates
7.2 Rate of evolution of WTI
7.3 The transregulatory domain
7.3.1 Evolution of homopolymer domains
7.3.2 Function of homopolymers
7.3.3 Conservation of the putative leucine zipper
7.4 A region of WTI with species specific function?
7.5 Selection of alternative splice sites
7.6 Evolution of the zinc fingers
7.7 Conservation of expression of the WTI gene in the intermediate mesoderm
7.7.1 Urinary system
7.7.2 Gonads
7.8 Additional sites of expression
7.9 WTI and limb development
7.10 Evolution of the vertebrate kidney and WTI
## Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Schematic representation of the human \textit{WTI} mRNA transcript</td>
<td>25</td>
</tr>
<tr>
<td>1.2</td>
<td>Secondary and tertiary structure of zinc fingers</td>
<td>27</td>
</tr>
<tr>
<td>1.3</td>
<td>Differentiation of the mammalian kidney</td>
<td>37</td>
</tr>
<tr>
<td>1.4</td>
<td>Differentiation of the metanephros</td>
<td>39</td>
</tr>
<tr>
<td>1.5</td>
<td>Nephron development</td>
<td>41</td>
</tr>
<tr>
<td>1.6</td>
<td>Secondary sexual determination</td>
<td>45</td>
</tr>
<tr>
<td>1.7</td>
<td>Phylogenetic tree of the vertebrates</td>
<td>58</td>
</tr>
<tr>
<td>1.8</td>
<td>The relationship between kidney development and evolution</td>
<td>59</td>
</tr>
<tr>
<td>1.9</td>
<td>The interrelationship between the gonad and urinary systems</td>
<td>61</td>
</tr>
<tr>
<td>1.1</td>
<td>Detection of sequences that hybridise to \textit{WTI} in a range of mammalian genomes</td>
<td>101</td>
</tr>
<tr>
<td>1.2</td>
<td>Detection of sequences that hybridise to \textit{WTI} in a range of vertebrate genomes</td>
<td>104</td>
</tr>
<tr>
<td>4.1</td>
<td>Predicted amino acid sequence of human \textit{WTI} showing the exon structure and positions of degenerate PCR primers used in this study</td>
<td>109</td>
</tr>
<tr>
<td>4.2</td>
<td>Testing of degenerate primers</td>
<td>113</td>
</tr>
<tr>
<td>4.3</td>
<td>Isolation of \textit{WTI} orthologs</td>
<td>115</td>
</tr>
<tr>
<td>4.4</td>
<td>Blunt ended cloning strategy used for cloning PCR products</td>
<td>116</td>
</tr>
<tr>
<td>4.5</td>
<td>Schematic representation of the human \textit{WTI} cDNA and the \textit{WTI} clones isolated by degenerate PCR</td>
<td>120</td>
</tr>
<tr>
<td>4.6</td>
<td>Isolation of chick and \textit{Xenopus WTI} orthologs</td>
<td>122</td>
</tr>
<tr>
<td>4.7</td>
<td>Isolation of zebrafish \textit{WTI}</td>
<td>124</td>
</tr>
<tr>
<td>4.8</td>
<td>Turbo cloning of PCR products</td>
<td>125</td>
</tr>
<tr>
<td>4.9</td>
<td>Comparison of human and zebrafish \textit{WTI} genomic structures in the zinc finger region</td>
<td>127</td>
</tr>
<tr>
<td>4.10</td>
<td>Diagram of \textit{WTI} clones isolated by library screening, in relation to the human \textit{WTI} gene</td>
<td>130</td>
</tr>
<tr>
<td>4.11</td>
<td>Isolation of ZF4 from chick</td>
<td>133</td>
</tr>
<tr>
<td>4.12</td>
<td>Genomic structure of zebrafish \textit{WTI}</td>
<td>137</td>
</tr>
<tr>
<td>5.1</td>
<td>Comparison of the predicted amino acid sequences of the \textit{WTI} orthologs</td>
<td>142</td>
</tr>
<tr>
<td>5.2</td>
<td>Multiple sequence alignment the of non coding region of exon 1 using \text{CLUSTAL V}</td>
<td>147</td>
</tr>
<tr>
<td>5.3</td>
<td>Investigation into the conservation of homopolymer domains in exon 1</td>
<td>150</td>
</tr>
<tr>
<td>5.4</td>
<td>Helical representation of the putative leucine zipper</td>
<td>152</td>
</tr>
<tr>
<td>5.5</td>
<td>Investigation into the conservation of the 17amino acid alternative splice\text{-exon 5}</td>
<td>155</td>
</tr>
<tr>
<td>5.6</td>
<td>Mutation, editing and conservation of exon 6</td>
<td>157</td>
</tr>
<tr>
<td>5.7</td>
<td>Schematic representation of the four zinc fingers of \textit{WTI}</td>
<td>159</td>
</tr>
<tr>
<td>5.8</td>
<td>Sequence of the mouse \textit{Wtl} cDNA clone KS/2</td>
<td>162</td>
</tr>
</tbody>
</table>
Figure 5.9 Graphs plotting $K_v$ against divergence time 169
Figure 5.10 Unrooted phylogenetic tree for exons 7-9 of $WTI$ 171
Figure 5.11 Graph of divergence of $WTI$ against divergence time 172

Figure 6.1 Detection of $WTI$ mRNA in pig, chick, and alligator 175
Figure 6.2 *Xenopus* $WTI$ mRNA expression 176
Figure 6.3 Expression of $WTI$ in the developing chick 180
Figure 6.4 Whole mount *in situ* hybridisation of alligator embryos-onset $WTI$ expression 184
Figure 6.5 $WTI$ expression in stage 9-11 alligator embryos 186
Figure 6.6 $WTI$ expression in stage 13-14 alligator embryos 188
Figure 6.7 *In situ* hybridisation of sections of the alligator mesonephros 196
Abbreviations

A adenosine
aa amino acid
ALL acute lymphoid leukemia
AML acute myeloid leukemia
AMV avian myeloblastosis virus
BCIP 5-bromo-4-chloro-3-indolyl-phosphate
bp base pair
BSA bovine serum albumin
BWS Beckwith Wiedemann Syndrome
C cytidine
cDNA complementary DNA
CML chronic myeloid leukemia
d days
DDS Denys Drash syndrome
DEPC diethylpyrocarbonate
dH2O distilled water
DIG digoxigenin
DMSO dimethylsulphoxide
DNA deoxyribonucleic acid
dNTP 2'-deoxyribosylnucleoside 5'-triphosphshates
dpc days post coitum
DTE dithioerythritol
DTT dithiothreitol
EDTA ethylenediaminetetra-acetate
EMBL european molecular biology laboratory
G guanosine
GCG Genetics Computer Group
(Glycine)5 GGGGG
GTF general transcription factor
GU genitourinary
HMG high mobility group
HPhillic hydrophilic
HPHobic hydrophobic
I inosine
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILNR</td>
<td>intralobar nephrogenic rest</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropylthiogalactose</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>$K_A$</td>
<td>number of nonsynonymous changes per nonsynonymous site</td>
</tr>
<tr>
<td>$K_S$</td>
<td>number of synonymous changes per synonymous site</td>
</tr>
<tr>
<td>KTS</td>
<td>lysine threonine serine</td>
</tr>
<tr>
<td>L Broth</td>
<td>Luria broth</td>
</tr>
<tr>
<td>LOH</td>
<td>loss of heterozygosity</td>
</tr>
<tr>
<td>MOPS</td>
<td>morpholinopropansulphonic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>Myr</td>
<td>millions of years</td>
</tr>
<tr>
<td>Myra</td>
<td>millions of years ago</td>
</tr>
<tr>
<td>NBT</td>
<td>nitroblue tetrazolium</td>
</tr>
<tr>
<td>N/A</td>
<td>not applicable</td>
</tr>
<tr>
<td>N/D</td>
<td>not done</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>p</td>
<td>short arm</td>
</tr>
<tr>
<td>PFA</td>
<td>paraformaldehyde</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>pfu</td>
<td>plaque forming units</td>
</tr>
<tr>
<td>PLNR</td>
<td>perilobar nephrogenic rests</td>
</tr>
<tr>
<td>(Proline)$_4$</td>
<td>PPPPP</td>
</tr>
<tr>
<td>(Proline)$_{13}$</td>
<td>PAPPAPPPPPPPPPPP</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>q</td>
<td>long arm</td>
</tr>
<tr>
<td>RACE</td>
<td>rapid amplification of PCR ends</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>rATP</td>
<td>5'-(pyro) adenosine triphosphate</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcriptase PCR</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SRO</td>
<td>smallest region of overlap</td>
</tr>
<tr>
<td>SSC</td>
<td>standard sodium citrate</td>
</tr>
<tr>
<td>St</td>
<td>stage</td>
</tr>
</tbody>
</table>
T thymidine
TCA trichloroacetic acid
TEMED N,N,N'N'-tetramethylethylenediamine
Tm melting temperature
Tris 2-amino-2-(hydroxymethyl)-1,3-propanediol
tRNA transfer RNA
U units of enzyme as defined by the manufacturer
UTP uridine 5'-triphosphate
UV ultraviolet
w weeks
WAGR Wilms' tumour aniridia genitourinary abnormalities mental retardation
WT Wilms' tumour
X-gal 5-bromo-4-chloro-3-indolyl-galactose
X-phosphate 5-bromo-4-chloro-3-indolyl-phosphate
ZF zinc finger
Chapter 1
Introduction
1.1 Introduction

One of the major achievements of molecular genetics so far has been to increase the understanding of the basis of human genetic disease. In recent years there has been a flood of information pinpointing genes in which lesions can lead to a particular disease. In many cases the reason why a genetic lesion is giving rise to a disease is easily understood. The classic examples being haemophilias A and B, which result from Factor VIII and Factor IX mutations, and the thalassemias which are due to mutations of the globin genes. In the vast majority of cases the gene was isolated using some information about the gene product - functional cloning. In the last few years genes have been isolated by virtue of their physical or genetic linkage to a disease locus i.e. by reverse genetics, otherwise known as positional cloning. These include the CFTR gene in cystic fibrosis, NFI in neurofibromatosis type 1, FMR in Fragile X syndrome and RB1 in retinoblastoma. It is then necessary to discover the function of the gene product and understand how malfunction can result in the particular disease phenotype.

Cancers and developmental abnormalities are two of the most complex disease types, the causes of which are just beginning to be unravelled through molecular genetics. Not infrequently these two superficially different classes of disease turn out to be intimately related. Cancers arise from a loosening of the strict growth controls on cells and development proceeds by the regulation of cell proliferation and differentiation. Wilms' tumour is an example of both a kidney developmental abnormality and a cancer.

A predisposition gene for Wilms' tumour, WT1, has been isolated by positional cloning on the basis of its genetic linkage to the disease. Since its isolation an important role for WT1 in kidney and gonad development, as well as in tumorigenesis, has been demonstrated. That it is not the sole gene involved in formation of Wilms' tumour is also clear. WT1 has been shown to have activities consistent with a role as a regulatory transcription factor but as yet its interactions in vitro and in vivo are not clearly understood. This particularly applies to the part it is playing in a mesenchyme to epithelial transition, and how WT1 mutants give rise to genitourinary abnormalities. Investigations have so far been hampered by the lack of a relevant model system.

An evolutionary comparison of WT1 will have many uses. A wider sequence comparison of the WT1 gene will facilitate the targeting and interpretation of structural, functional and mutational analysis. The comparison between the mouse and human sequences reveals very little due to the almost complete similarity between
the two genes. At the same time the isolation of \( WTI \) orthologs will lay the foundations for a series of animal models in which \( WTI \) can be investigated and manipulated. In the case of \( WTI \) this is a particularly appropriate line of investigation because vertebrate kidney development is based on an initial plan which is then modified in the different classes. This seems to have occurred through modifications at the later stages of development so that the lower vertebrate kidneys have resemblances to foetal mammalian kidneys. By looking at lower vertebrate species it may be possible to elucidate the role of \( WTI \) in early human kidney development. The investigation of \( WTI \) may also increase the understanding of the evolution of the vertebrate kidney.

1.2 Wilms' Tumour and the identification of a putative predisposition gene-\( WTI \)

1.2.1 Incidence of Wilms' tumour

Wilms' tumour or nephroblastoma is an embryonic cancer of early childhood with 90% of cases presenting in the first 7 years of life, affecting 1 in 10,000 children (Young and Miller 1975). The early age of onset and the often undifferentiated histopathology of the tumour has suggested that Wilms' tumour can be a result of an error in the kidney developmental process, and it is thought to arise from the nephrogenic cells of the kidney (Nicholson 1931, Mierau et al, 1987). A comprehensive study of the incidence of Wilms' tumours has been produced by the US National Wilms' tumour Survey (Beckwith 1983, Breslow et al., 1988, Olson et al 1993). Although the tumour is observed world-wide, the incidence varies between ethnic groups: it is higher in the American Black population than Caucasian (Kramer 1984) and lower in Asian than European populations (Tadokoro et al, 1992). Most cases are sporadic, but 1% of cases are familial with tumour predisposition inherited as an autosomal dominant trait with variable penetrance (Matsunga 1981). In addition incidence of associated congenital abnormalities in WT patients is greater than in the general population. Up to 15% of Wilms' tumour patients have associated congenital abnormalities. In particular, 1% have aniridia, 4% have genitourinary abnormalities, 1.5% have cardiopulmonary abnormalities and 3% have hemihypertrophy (Breslow and Beckwith 1982, Olson et al., 1993). In males who have Wilms' tumour and aniridia the frequency of the genitourinary abnormalities rises to 40-50% (Shannon et al., 1982, Turleau et al., 1984). The abnormalities can be grouped into several syndromes: WAGR, DDS, Perlman and BWS (Table 1.1). In the case of WAGR patients large cytogenetic deletions have been observed.
Table 1.1 Congenital abnormality syndromes associated with Wilms' tumour

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Abbreviation</th>
<th>Phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAGR</td>
<td>WAGR</td>
<td>Wilms' tumour, Aniridia, *Genitourinary Abnormalities, Mental Retardation</td>
<td>Miller et al., 1964</td>
</tr>
<tr>
<td>Denys-Drash</td>
<td>DDS</td>
<td>Wilms' tumour, Gomerular Nephropathy, **Genitourinary Abnormalities</td>
<td>Denys et al., 1967</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Drash et al., 1970</td>
</tr>
<tr>
<td>Beckwith Wiedemann</td>
<td>BWS</td>
<td>Wilms' Tumour, Adrenocortical Carcinoma, Hepatoblastoma, Foetal Gigantism,</td>
<td>Beckwith 1969</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exomphalos, Macroglossia, Organomegaly, Hemihypertrophy, Gonadal Abnormalities</td>
<td>Wiedemann 1964</td>
</tr>
<tr>
<td>Perlman</td>
<td></td>
<td>Wilms' Tumour, Foetal Gigantism, Bilateral Nephromegaly,</td>
<td>Perlman et al., 1973</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nephroblastomatosis, Cryptorchidism</td>
<td></td>
</tr>
</tbody>
</table>

*Genitourinary Abnormalities: Cryptorchidism, Hypospadias, Fused Kidneys, Double Collecting System

**Genitourinary Abnormalities: A spectrum of abnormalities, from the abnormalities above all the way to Pseudohermaphroditism
1.2.2 Knudson and Strong two hit model for tumorigenesis.

Knudson (1971) originally proposed his two hit model of tumorigenesis for retinoblastoma (RB), a childhood eye tumour. Using statistical analysis he showed that the incidence of unilateral (in one eye), bilateral (in both eyes), familial and sporadic cases fitted a minimum of two mutations or 'hits' occurring to initiate tumorigenesis. He proposed that for familial cases the first mutation is present in the germline and so only one somatic mutation had to occur. In non germline cases both mutations had to be somatic and so the later onset than in familial cases was explained by the longer time for two mutations to occur. The onset of bilateral cases tended to be earlier, in the same range as familial tumours, also indicating the presence of a germline predisposition mutation. The tumour is limited to childhood because the target cell population are embryonic retinal precursor cells which disappear with time after birth. That the mutations resulted in the loss of a tumour suppressor was suggested by the association of retinoblastoma with deletions at chromosome 13q14 as well as loss of heterozygosity for this region in the tumours, the latter being consistent with a second hit (which would effectively expose the first hit in a recessive gene) (Cavenee et al., 1983).

That the malignant phenotype can be due to the loss of gene function as opposed to dominantly acting oncogenes had been suggested by experiments which had demonstrated that a wild type cell could suppress the tumorigenic effect of malignant cells on cell fusion (Harris et al., 1969, Stanbridge 1976). The tumour suppressor gene involved in retinoblastoma, $RB1$, has been cloned (Friend et al., 1986, Lee et al., 1987) and so far all cases have been found to be consistent with Knudson's model (Goodrich and Lee 1990).

Comparison of the incidence of Wilms' tumour with retinoblastoma showed many similarities, an earlier onset of the presumed germline and bilateral cases, and in the few familial cases autosomal dominant inheritance. On the basis of this it was proposed that Wilms' tumorigenesis could also arise from two 'hits' (Knudson and Strong 1972). The major difference between the two was that although the sporadic bilateral (i.e. germline) cases of both were frequent the number of familial cases was much lower in WT. These WT familial cases did not segregate with chromosome 11p to which the sporadic cases had been mapped. This indicated that although there were similarities in the molecular basis of the two childhood tumour types, WT maybe more complicated, with more than locus involved.
1.2.3 Mapping of the WAGR locus to 11p13

The first information concerning the location of a predisposition gene came from the karyotypic analysis of rare cases where Wilms' tumour was associated with congenital abnormalities and visible cytogenetic deletions. An interstitial deletion in the short arm of chromosome 11 was found in several cases of WAGR syndrome (Franke et al., 1977). The localisation was refined in two cases where the only deleted region was 11p13 (Franke et al., 1979, Riccardi et al., 1978). This would be consistent with the first hit of Knudson theory. That 11p may also be involved in second hit was shown by loss of heterozygosity for this region in tumours compared to somatic tissue (Fearon et al., 1984, Orkin et al., 1984, Reeve et al., 1984). Loss of heterozygosity on 11p was also found to occur in other types of tumours indicating a more general tumour suppressing activity (Koufos et al., 1985). That there was a tumour suppressing activity present on chromosome 11 was demonstrated experimentally by suppressing tumorigenicity of a putative Wilms' tumour cell line in nude mice by introducing chromosome 11 (Weissman et al., 1987).

1.2.4 Isolation of WTI

More WAGR cases containing constitutional deletions of 11p13 were used to define the WAGR region to between the catalase and follicle stimulating hormone β (FSHβ) genes (van Heyningen et al., 1985, Glaser et al., 1986, Porteous et al., 1987, Bickmore et al., 1989). To define the exact extent of the deletions detailed physical and genetic maps were built up using newly isolated markers for this region (Porteous et al., 1987, Gessler et al., 1989, Compton et al., 1990). Rare cutting enzymes and pulsed-field gel electrophoresis were used to identify CpG islands within the region (Bickmore et al., 1989) as they are often located at the 5' end of genes (Bird 1986). A very detailed physical map of the WAGR region was produced using irradiation reduced hybrids in which the only human component was a small part of chromosome 11p (Rose et al., 1990).

Lewis et al., (1988) were the first to demonstrate the homozygous deletion of an 11p13 marker in a sporadic Wilms' tumour (WiT-3) in a patient with no associated abnormalities. The smallest region of overlap (SRO) between the deletions in WiT-3 was defined, allowing the Wilms' tumour predisposition region to be limited to 345kb (Rose et al., 1990). Two single copy probes isolated from chromosome 11 cosmids were found to map to this region and were used to isolate a partial cDNA from a pre-B cell library that contained an open reading frame (WT33). This was found to hybridise to an RNA transcript in mouse and baboon kidney (Call et al., 1990).
Figure 1.1 Schematic representation of the human WTI mRNA transcript

The exon structure is shown including the two alternative splices (17aa-exon5 and KTS) (Haber et al., 1991, Gessler et al., 1992). The major features of the encoded protein, the four zinc fingers, the putative leucine zipper, and the proline/glutamine rich transcription regulatory domain including the glycine and proline rich domains within it, are shown.
same gene, called *WTI*, was also isolated by chromosome jumping between CpG islands (Gessler et al., 1990).

### 1.3 Structure and function of *WTI*

The cDNA isolated by Gessler et al., (1990) was 3kb long, corresponding to the size of transcript seen in Northern analysis, and contained a 1,725bp open reading frame (Figure 1.1). Subsequent analysis of the genomic sequence has shown that the transcript is made up of 10 exons, covering 50kb (Haber et al., 1991, Gessler et al., 1992).

#### 1.3.1 Zinc fingers

The sequence of the C-terminal end of the predicted protein was found to correspond to four zinc finger motifs of the Cys2 His2 type, first identified in *Xenopus TFIIIA* (Miller et al., 1985). Zinc fingers contain conserved cysteine and histidine residues which chelate a zinc ion holding together the tertiary structure of the finger (Miller et al., 1985, Pavletich and Pabo 1991, Figure 1.2a). Over 1000 finger motifs have been identified in more than two hundred proteins, mostly transcription factors e.g. *GLI* family (Ruppert et al., 1988), and *Sp1* (Kadonaga et al., 1987). These motifs have been shown many times to have sequence specific DNA binding activities (reviewed in El-Baradi and Pieler 1991) and the orientation of three finger regions, *Zif*268 (*EGRI*) (Pavletich and Pabo 1991), *GLI1* (Pavletich and Pabo 1993) and *tamtrack* (Fairall et al., 1993) when bound to DNA has been elucidated by X-ray crystallography of DNA-protein co-crystals (Figure 1.2b). The *WTI* fingers 2-4 most closely resemble the three fingers in *EGR2* (*Krox 20*) (Joseph et al., 1988) with an identity at the amino acid level of 67% and a similarity of 75% (Figure 1.2c). No sequence similarity is seen outside the zinc fingers except for a similar richness in certain amino acids. *EGR2* belongs to the *EGR* family which contain virtually identical zinc fingers but with only a few regions of homology in the rest of the gene (Crosby et al., 1992). They have a conserved function in the early growth response to serum stimulation (Sukhatme 1992) and development (Nieto et al., 1991).

#### 1.3.2 Proline/glutamine rich region

Upstream of the zinc fingers the *WTI* coding sequence is particularly rich in proline and glutamine residues including a run of fifteen amino acids, thirteen of which
Figure 1.2 Secondary and tertiary structure of zinc fingers

a NMR determined structure for the second zinc finger of SWI5. This shows the antiparallel β strands followed by the α helix which are held in the 'finger' conformation by the chelation of a zinc$^{2+}$ ion by two cysteine and two histidine residues and the hydrophobic core (Neuhaus et al., 1990).

b The orientation of the three zinc fingers of EGR1 in relationship to a DNA binding site as determined by the X-ray analysis of the protein-DNA co-crystal (Pavletich and Pabo 1991). The finger contacts the DNA in the major groove with the 'tip' of the finger closest to DNA, i.e. the C-terminal end of the β strand and N-terminal end of the α helix. The fingers are also shown to be wrapping around the DNA. (Illustrations taken from Travers 1993)

c Comparison of the amino acid sequence of the zinc fingers of EGR1&2 with fingers 2-4 of WT1. The conserved zinc chelating residues are shown in bold and the structural regions of the fingers indicated, including the regions involved in DNA binding. The residues directly involved in base recognition are underlined.

d Pattern of contacts of the three zinc fingers of EGR1 (Zif 268) with their triplet base recognition sequence. The pattern is identical for fingers 1&3 but is different in ZF2 due to the substitution of an arginine for a threonine residue. The positions within the zinc fingers involved in base contacts are circled and the α helix is shaded. *Indicates an aspartate residue which interacts with the first arginine to stabilise its interaction with the DNA. (Illustration from Fairall et al., 1993)
Zinc Finger Consensus: $XF/\text{YXC} \times 2/4 \text{CXXXFXXXXXLXXH}X3/4 \text{HTGEK}$

**WT1 ZF2**
PYQCDFKCERRFSRSDQLKBRHQRRTHTGVK

**EGR2 ZF1**
PYPCAPACDRRFSRSDLTKHTHTHHK

**Egr1 ZF1**
PYACPVSCDRRFSRSDLTKHTHTGHQK

**WTI ZF3**
PFQC..KTCQRKFSRSDHLKHTRTHTGKTSEK

**EGR2 ZF2**
PFQC..RICMRNFSRSDHLLTHRHTG..EK

**Egr1 ZF2**
PFQC..RICMRNFSRSDHLLTHRHTG..EK

**WTI ZF4**
PFSCRWPSCKKKFARSDELVPHHNNH

**EGR2 ZF3**
PFACDY..CGKFKARSERKHTKIH

**Egr1 ZF3**
PFACDI..CGKFKARSERKHTKIH

β sheet → α helix → linker → DNA binding

---

d

**Zif268 F1 and F3**

**Zif268 F2**
are proline. Although there is no actual sequence homology between them, regions rich in proline, glutamine or acidic amino acids have been found in an increasing number of transcription factors, OTF2 (proline and glutamine rich), Sp1 (glutamine rich), EGR family (proline rich) and CTF/NF1 (proline rich) (Mitchell and Tjian 1989). Transcription regulatory activity has been localised to this region in CTF/NF1 (Mermond et al., 1988). The mechanism by which these domains act to alter transcription is just beginning to be elucidated. In higher eukaryotes transcription is carried out by RNA polymerase II in association with many general transcription factors (GTF) e.g. TFIIA-I (Zawel and Reinberg, 1993). Recently interactions between these GTFs and regulatory transcription factors have been described e.g. between c-rel and TFIIr—the TATA box binding protein (Kerr et al., 1993).

1.3.3 Alternative splices

WT1 is subject to alternative splicing of the RNA transcript at two positions (Figure 1.1). One is an insertion of 3 amino acids, KTS (lysine, threonine, serine) in the linker region between zinc fingers 3 and 4 (Haber et al., 1990) and the other occurs between the proline/glutamine rich region and the fingers inserting an extra 17 amino acids (Haber et al., 1991). From characterisation of the genomic structure of WT1 it is known that the 3 amino acid insertion is achieved through the use of alternative splice donor sites in exon 9, and the 51 nucleotides which specify the 17 amino acid splice are contained within a separate exon, exon 5 (Haber et al., 1991, Gessler et al., 1992). The in vivo function of these splices is not known. The 17 amino acid insertion is at the C-terminal end of a putative leucine zipper (Landshultz et al., 1988). Since zippers are dimerisation motifs (Kouzandes et al., 1988) insertion at its end might change dimerisation characteristics of the protein.

The KTS insertion is in a finger linker region, which is a highly conserved sequence of amino acids in the TFIIIA type fingers. Insertions and changes in this sequence have been shown to change the DNA binding characteristics of TFIIIA (Choo and Klug 1993). The mRNA transcripts containing the KTS insertion have been observed to be the most plentiful RNA transcript (Brenner et al., 1992).

In contrast to many genes a tissue specific pattern of expression of the four possible forms of the RNA has not been observed. Instead in the tissues examined a constant ratio is present (Haber et al., 1991). On cloning the mouse homolog of WT1 (Buckler et al., 1991), which is 95% conserved at the amino acid level, both
alternative splices were also seen and similar ratios were observed in a wide range of tissues (Haber et al., 1991).

1.3.4 Identification of the WTI protein

Using WTI specific antibodies, WTI has been shown to be a 49-54kDa protein, this correlates well with the size predicted from the primary sequence of 47-48.9kDa (Morris et al., 1991, Telerman et al., 1992). The range in size is thought to arise from alternative splicing or to phosphorylation. The latter could not be detected when transiently transfected COS-1 cells were 32P labelled although there are potential phosphorylation sites (Morris et al., 1991). Subcellular localisation to the nucleus consistent with WTI being a transcription factor has been shown by cellular fractionation, immunohistochemistry (Morris et al., 1991, Telerman et al., 1992) and epitope tagging (Pelletier et al., 1991c).

1.3.5 Transcription factor activity

1.3.5.1 DNA binding

Numerous studies have tried to define in vitro, in various ways, the consensus DNA binding sites for the WTI zinc fingers. The -KTS but not the +KTS isoform was found to bind sequences extremely similar to and including the EGR1 consensus (GCGGGGGGCG) but not to other GC rich sequences, such as those bound by Sp1 and myb (Rauscher et al., 1990). It is not surprising that WTI is capable of binding the EGR1 consensus considering the similarity of fingers 2-4 to those of EGR1, especially in those regions known to interact with the DNA (Figure 1.2c&d). The +KTS isoform has also been shown to bind DNA in a sequence specific manner including some of the -KTS binding sites (Bickmore et al., 1992, Wang et al., 1993b). As yet no true consensus sequence has been identified for either isoform.

1.3.5.2 Target promoters

WTI has been shown in transfection systems to bind and down-regulate transcription from genes whose protein products are known to have mitogenic properties, EGR1 (Madden et al., 1991), IGF2 (Drummond et al., 1992), PDGFA chain (Gashler et al., 1992, Wang et al., 1992) and IGF1R (Werner et al., 1993).
These genes could well have a physiological role in kidney development. *EGR1* is expressed in response to serum stimulation of cells, *WT1* may antagonise this mitogenic response to growth factors by repressing *EGR1* expression or repressing expression of *EGR1* activated genes, as it binds to *EGR1*-like binding sites. This may only occur in certain circumstances because unlike *EGR1*, *WT1* is not expressed in response to serum stimulation (Morris et al., 1991). *IGF2* (insulin-like growth factor 2) is a growth factor which is thought to have an important role in development. The overexpression of *IGF2* mRNA in Wilms' tumours (Reeve et al., 1985, Scott et al., 1985) makes it a good candidate for regulation by *WT1*. Down-regulation of *IGF2* gene expression can be seen in the kidney as *WT1* is up-regulated (Drummond et al., 1992, Yun et al., 1993). *IGF1R* (insulin-like growth factor 1 receptor) is a transmembrane tyrosine kinase thought to mediate the effects of IGF1&2. In Wilms' tumours there is an inverse relationship between the levels of *WT1* and *IGF1R* expression which correlate with histological differences (Werner et al., 1993). Blockage of *IGF1R* in nude mice using antibodies can prevent growth of Wilms' tumours after inoculation with transformed cells (Gansler et al., 1989). *PDGFA* (platelet derived growth factor A chain) is another growth factor and is a powerful chemoattractant and mitogen for cells of mesenchymal origin (Deuel 1987). The importance of *PDGFA* in kidney differentiation has yet to be demonstrated but mesenchymal cells play a large role in kidney formation (see below). Therefore if functional *WT1* is lost in Wilms' tumours the loss of repression of these genes and therefore their expression may lead to a stimulation of growth as opposed to differentiation (Sukhatme 1992).

### 1.3.5.3 Mapping of functional domains

It has been possible, for some transcription factors, to define regions responsible for the transacting activity by making deletions in these regions. It has been shown that *EGR1* contains four regions which are responsible for transactivation and one that mediates transrepression (Russo et al., 1993). Binding of *WT1* to DNA is necessary for transrepression and activation. The transrepression activity has been mapped to the proline/glutamine rich region as transferring this region to a heterologous DNA binding motif (*GAL4*) also transfers the transrepression activity (Madden et al., 1993). Finer mapping of this region has failed to produce any convincing further localisation of function (Madden et al., 1993, Wang et al., 1993a). The deletion of three homopolymer motifs, (Glycine)$_5$ (GGGGG), (Proline)$_{13}$ (PAPPPAPPPPPPPPP), (Proline)$_4$ (PPPP) and mutations in the putative leucine
zipper reduce but do not ablate transrepression activity suggesting that the effect is due to a structural disturbance rather than function residing in this region (Figure 1.1).

1.3.5.4 Protein-protein interactions and modulation of WTI activity

From the initial simple model of transcriptional regulation in prokaryotes it has become clear that eukaryotic transcriptional regulation occurs through the interaction of multiple factors often bound to multiple promoter sequences (Zawel and Reinberg, 1993). There is some evidence for the modulation of WTI action by association with another tumour suppressor, p53 (Maheswaran et al., 1993). Coimmunoprecipitations have shown the presence of p53 and WTI in the same complex. Wild type WTI (-KTS) could enhance the activation by wild type p53 of one of its targets; MCK (muscle creatine kinase) promoter. Also the suppression of transcription using EGR1 consensus sites by WTI (-KTS) was found only to occur in the presence of wild type p53, showing that transcriptional repression mediated by WTI is not an intrinsic property, but modulated by the cellular environment. The most often observed effect on transcription by WTI is transrepression but this is dependent on the number and positioning of binding sites and on cell type. The need for multiple binding sites for repression to occur may be due to WTI interacting with itself as has been shown for Sp1 (Su et al., 1991).

1.3.5.5 Regulation of WTI expression

As yet very little is known about what controls the expression of WTI. It may be that other WT predisposition genes are located upstream of WTI and are involved in its regulation. The region 5' to the transcriptional start position has been cloned in both mouse and human (Pelletier et al., 1991c, Campbell et al., 1993, Hofmann et al., 1993). Similarities exist between these regions in that neither possess TATA or CCAAT elements but there are numerous putative Sp1, EGR1, WTI, PAX8, PAX2, AP2, AP4 and GAGA binding sites. The presence of numerous Sp1 sites is consistent with other TATA-less, GC-rich promoters. The binding of Sp1 to the promoter region has been demonstrated in vitro (Hofmann et al., 1993). From the sequence of the WTI promoter it was predicted that WTI may bind its own promoter (Hofmann et al., 1993). This has now been shown to be the case in vitro (Campbell et al., 1993). All four isoforms of WTI can down-regulate its expression, the isoform which most efficiently does this was the +17aa, +KTS isoform (Rupprecht et al., 1994). A basal promoter region has also been defined and similar to other GC-rich TATA-less
promoters it is non-cell-specific. It has been proposed that in at least one cell line an enhancer of transcription lies 50kb downstream in exon 10 (Frazier et al., 1994).

1.4 Evidence that WTI is a predisposition gene for Wilms' tumour

On the basis of the structure of the predicted protein it was proposed that WTI could be acting as a transcription factor involved in the control of kidney development and that loss of the gene function would perturb development giving cells the opportunity to become neoplastic. In order to demonstrate that WTI is involved in the development of Wilms' tumours the presence of intragenic mutations, the ultimate test for a candidate gene, in tumour DNA was investigated. In addition the expression pattern was investigated to see if it was consistent with a role in kidney development.

1.4.1 Mapping deletions in Wilms' tumours to the WTI locus

Very few Wilms' tumours have homozygous deletions in WTI detectable by Southern blotting (Cowell et al., 1991, Royer-Pakora et al., 1991, Ton et al., 1991a). In fact most of the Wilms' tumours examined expressed WTI, indicating that in most cases more subtle lesions must be occurring than complete ablation of the gene product (Gessler et al., 1990, Huang et al., 1990, Pritchard-Jones et al., 1990). Comparison of two tumours WiT-13 and PER which are both homozygously deleted for the WT region defined a region of over 150kb (Gessler et al., 1990, Rose et al., 1990). As the approximate genomic region covered by a gene is 50kb it was possible that WTI was just very closely linked to the real predisposition gene. A second mRNA transcript had already been isolated within the WT region which was also expressed in the developing kidney (Bonetta et al., 1990, Huang et al., 1990). The WT region was reduced by analysis of another sporadic Wilms' tumour homozygously deleted for the WT region (Ton et al., 1991a). The deletions were found not to extend 3' of the first few exons of WTI and with the 5' end of the WT region defined by the 5' extent of the deletion in PER, the WT region was reduced to 16kb. The smallest region of overlap of the homozygous deletions in a sporadic Wilms' tumour was found to extend 3' of the WTI gene deleting only the last zinc finger and not extending as far as the next CpG island (Cowell et al., 1991). These last two deletions do not overlap but both delete part of the WTI gene: this heavily implicates WTI as the predisposition gene.
1.4.2 Intragenic deletions

In a sporadic unilateral Wilms' tumour, a small intragenic deletion was found, the deletion of 25bp at an exon/intron junction resulted in the deletion of zinc finger 3, so altering or abolishing DNA binding activity (Haber et al., 1990, Rauscher et al., 1990, Wang et al., 1993a). LOH on 11p was detected but the tumour was heterozygous for the WTI deletion, and so had one normal copy of the gene, suggesting that WTI mutation was the second hit and the predisposing mutation was due to LOH for 11p.

The first reported case where a tumour became homozygous for an intragenic deletion in WTI was in a patient with bilateral tumours, with no associated abnormalities, who had a constitutional deletion of <11kb within the WTI gene. This caused a 97bp deletion in the RNA transcript and truncation of the protein by an in-frame stop codon. Both tumours were shown to have become homozygous for this deletion of WTI but analysis of chromosome 11 markers identified these as separate events (Huff et al., 1991). A second phenotypically normal case has been shown to have a constitutional deletion of two exons that became homozygous in the tumour (Tadekoro et al., 1992) (Table 1.2).

Analysis of the remaining allele in four unilateral WAGR patients with constitutional deletions spanning the WT locus showed that in these cases the tumours contained mutations in the WTI gene that would produce a truncated protein product (Brown et al., 1992, Baird et al., 1992a, Gessler et al., 1993) (Table 1.3). A second independent hit would be expected in these cases as becoming homozygous for these large deletions would probably lead to cell lethality. These cases therefore show that in most instances Knudson's two hit model does apply for Wilms' tumour and WTI.

So far WTI mutations have been detected in only about 10% of the Wilms' tumours analysed, and most have been limited to the zinc fingers (Tables 1.2-1.4). These will be expected to change or ablate the DNA binding activity. There have been a few reports of changes in the rest of the protein. A point mutation has been found in a Wilms' tumour that changes a glycine to aspartic acid in exon 3 (codon 201) (Park et al., 1993c); this has been shown to change WTI from a transrepressor to transactivator. The removal of exon 2 by alternative splicing has also been shown to change the properties of WTI in the same way. This exclusion of exon 2 has been found only in cells derived from Wilms' tumours and in the absence of other WTI mutations. The resultant protein is found to be unable to repress growth of cells unlike wild type WTI, implicating it in tumorigenesis (Haber et al., 1993). Another
Table 1.2 WTI mutations in WT with no associated congenital abnormalities

Summary of characterised WTI mutations in Wilms' tumour cases where there are no associated congenital abnormalities. The genetic events which are thought to be the first and second hits in tumorigenesis are shown.

<table>
<thead>
<tr>
<th>Tumour Incidence</th>
<th>First mutation</th>
<th>Karyotype</th>
<th>Mutation in WTI gene</th>
<th>Change in WTI protein</th>
<th>Status of mutation in tumour</th>
<th>First hit</th>
<th>Second hit</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>unilateral tumour</td>
<td>N/D</td>
<td>25bp deletion at exon 9/ intron 9 boundary</td>
<td>Exon skipping deletes ZF3</td>
<td>heterozygous</td>
<td>LOH 11</td>
<td>WTI mutation</td>
<td>Haber et al., 1990</td>
<td></td>
</tr>
<tr>
<td>unilateral tumour</td>
<td>N/D</td>
<td>missense mutation in exon 8</td>
<td>366R-C in ZF2</td>
<td>heterozygous</td>
<td>?</td>
<td>?</td>
<td>Little et al., 1992a</td>
<td></td>
</tr>
<tr>
<td>unilateral tumour</td>
<td>46XX</td>
<td>4bp insertion in exon 2</td>
<td>truncation in ZF2</td>
<td>homozygous</td>
<td>WT1 truncation</td>
<td>LOH 11p</td>
<td>Park et al., 1993b</td>
<td></td>
</tr>
<tr>
<td>bilateral germline</td>
<td>46XX</td>
<td>&lt;11kb deletion from intron 5 to 6</td>
<td>exon 6 deleted, truncation in exon 7</td>
<td>1) homozygous 2) homozygous</td>
<td>1) WTI truncation 2) WTI truncation</td>
<td>1) LOH 11p 2) LOH 11p</td>
<td>Huff et al., 1991</td>
<td></td>
</tr>
<tr>
<td>bilateral germline</td>
<td>N/D</td>
<td>chain termination mutation in exon 9</td>
<td>390R-STOP, truncation in ZF3</td>
<td>1) homozygous 2) homozygous</td>
<td>1) WTI truncation 2) WTI truncation</td>
<td>1) LOH 11p 2) ?</td>
<td>Little et al., 1992a</td>
<td></td>
</tr>
<tr>
<td>unilateral germline</td>
<td>46XY</td>
<td>8kb deletion introns 5-7</td>
<td>truncation in ZF3</td>
<td>homozygous</td>
<td>WT1 truncation</td>
<td>LOH 11p</td>
<td>Tadokoro et al., 1992</td>
<td></td>
</tr>
<tr>
<td>unilateral tumour</td>
<td>46XY</td>
<td>1bp insertion in exon 10</td>
<td>frameshift in ZF4, extension of 66aa</td>
<td>homozygous</td>
<td>WT1 mutation</td>
<td>LOH</td>
<td>Coppes et al., 1993a</td>
<td></td>
</tr>
<tr>
<td>unilateral tumour</td>
<td>46XX</td>
<td>1bp deletion in exon 10</td>
<td>frameshift in ZF4, extension of 66aa</td>
<td>homozygous</td>
<td>WT1 mutation</td>
<td>LOH</td>
<td>Coppes et al., 1993a</td>
<td></td>
</tr>
<tr>
<td>unilateral germline</td>
<td>46XX</td>
<td>chain termination mutation in exon 8</td>
<td>362R-STOP truncation in ZF2</td>
<td>homozygous</td>
<td>WT1 truncation</td>
<td>LOH</td>
<td>Coppes et al., 1993a</td>
<td></td>
</tr>
<tr>
<td>unilateral germline</td>
<td>46XX</td>
<td>missense mutation in exon 9</td>
<td>394R-W</td>
<td>homozygous</td>
<td>WT1 mutation</td>
<td>LOH</td>
<td>Akasaka et al., 1993</td>
<td></td>
</tr>
<tr>
<td>unilateral tumour</td>
<td>N/D</td>
<td>chain termination mutation in exon 9</td>
<td>390R-STOP truncation in ZF3</td>
<td>homozygous</td>
<td>WT1 truncation</td>
<td>LOH</td>
<td>Varanasi et al., 1994</td>
<td></td>
</tr>
<tr>
<td>unilateral tumour</td>
<td>N/D</td>
<td>deletion &amp; insertion in exon 9</td>
<td>truncation in ZF3</td>
<td>homozygous</td>
<td>WT1 truncation</td>
<td>LOH</td>
<td>Varanasi et al., 1994</td>
<td></td>
</tr>
<tr>
<td>unilateral tumour</td>
<td>N/D</td>
<td>1) chain termination mutation in exon 8 2) 7bp insertion in exon 3</td>
<td>1) 362R-STOP truncation in ZF2 2) frameshift and truncation</td>
<td>heterozygous</td>
<td>WTI mutation</td>
<td>WTI mutation</td>
<td>Varanasi et al., 1994</td>
<td></td>
</tr>
<tr>
<td>unilateral tumour</td>
<td>N/D</td>
<td>chain termination mutation in exon 8</td>
<td>362R-STOP truncation in ZF2</td>
<td>homozygous</td>
<td>WT1 truncation</td>
<td>LOH</td>
<td>Varanasi et al., 1994</td>
<td></td>
</tr>
<tr>
<td>unilateral tumour</td>
<td>N/D</td>
<td>deletion &amp; insertion in exon 8</td>
<td>frameshift and truncation</td>
<td>heterozygous</td>
<td>?</td>
<td>?</td>
<td>Varanasi et al., 1994</td>
<td></td>
</tr>
</tbody>
</table>
point mutation in a Wilms' tumour associated with BWS has been reported. This is a phenylalanine to serine change in exon 2 (codon 154), as yet it has to be shown if this results in a change in the activity of \textit{WTI} (Park \textit{et al.}, 1993b).

1.5 Kidney development and \textit{WTI}

In order to demonstrate how \textit{WTI} could be involved in Wilms' tumour formation, the expression of \textit{WTI} was investigated in the developing human embryo and Wilms' tumours (Pritchard-Jones \textit{et al.}, 1990). Subsequently similar patterns of developmental expression have been seen in mouse and rat (Buckler \textit{et al.}, 1991, Armstrong \textit{et al.}, 1992, Sharma \textit{et al.}, 1992).

1.5.1 The developing kidney

The development of the mammalian metanephric kidney has been investigated extensively as an example of how development occurs as a result of the interaction between the epithelial and mesenchymal components. The metanephric kidney is particularly amenable to investigation as it can be cultured and manipulated \textit{in vitro}. It has been shown using transfilter culturing techniques that differentiation is dependent on a series of reciprocal inductions between these two components (Saxen 1970).

Kidneys arise from the nephrogenic mesenchyme of the intermediate mesoderm which lies between the lateral plate mesoderm and the somitic mesoderm and abuts onto the coelom (Gilbert 1991). A wave of differentiation extends caudally down the mesoderm from the region of the fourth somite forming the filtration units, the nephrons (Saxen 1987). Three kidneys are formed successively, the pronephros, mesonephros, and metanephros (Figure 1.3). A wave of degeneration follows the wave of differentiation removing the first two kidneys during foetal life leaving the metanephros as the functional adult kidney. Although the three kidneys are referred to as separate entities they are very closely associated and often it is hard to distinguish the very transitory pronephros from the beginning of mesonephric differentiation (Torrey 1965).

Kidney development is initiated when the pronephric analage form in the intermediate mesoderm, at about the level of the 4th somite, which then differentiate into the pronephric tubules. The pronephric duct extends caudally from this region until it joins the cloaca. As it extends it induces the differentiation of more pronephric
Figure 1.3 Differentiation of the mammalian kidney

a The first pronephric tubules are seen in human at about day 22 at the level of the 4th somite on either side of the midline. The nephric (pronephric) duct extends caudally inducing nephron differentiation.

b The first mesonephric differentiation occurs about day 24 in human. At this time the pronephric tubules are already regressing and the nephric (mesonephric) duct has reached the cloaca.

c The metanephros starts to differentiate about the 5th week of gestation, induced by the formation of the ureteric bud from the nephric duct near to the cloaca. The pronephros has completely regressed and the mesonephros is starting to do so. The gonad is differentiating on the medial side of the mesonephros. (Illustrations from Gilbert 1991 after Saxen 1987)
tubules and then the mesonephric tubules from the nephrogenic mesoderm. In the region of the mesonephros it is known as the mesonephric duct (Figure 1.3b).

The initiation of the metanephros occurs when the ureteric bud branches off the mesonephric duct just above the cloaca (Figure 1.3c). This bud induces the condensation of the surrounding nephrogenic mesenchyme (Grobstein 1955, Sariola et al., 1989, Bard and Woolf, 1992). The bud then divides dichotomously to form many branches (Figure 1.4a). As this is happening the mesenchyme proliferates and follows one of three pathways either leading to apoptosis (Coles et al., 1993), formation of the supporting stroma or differentiation into epithelia of the nephrons.

Northern blotting and in situ hybridisation of foetal metanephric kidneys shows the up-regulation of WTI mRNA as the metanephric blastema is induced to condense by the ureteric bud. The induction leads to the formation of characteristic comma shaped and then S-shaped bodies as the nephric tubules differentiate (Figure 1.5a). The tubules further differentiate into Bowman's capsules, proximal, and distal tubules, including loops of Henle. The distal end joins up to the collecting duct system formed from the branched ureteric bud. As differentiation proceeds, WTI expression is limited to the developing nephrons becoming localised to the Bowman's capsule, and then to the podocytes (Pritchard-Jones et al., 1990). No WTI expression has been seen in the ureteric bud or the collecting system. There is a second wave of differentiation of the nephrogenic blastema when the collecting ducts are no longer branching but are still competent to induce the metanephric blastema, forming arcades of nephrons which empty into the same collecting duct, thus greatly increasing the capacity of the kidney. In humans, mice, and rats the level of WTI expression declines after the last wave of nephron differentiation and loss of the renal blastema has occurred either before birth or neonatally (Pritchard-Jones et al., 1990, Buckler et al., 1991, Armstrong et al., 1992, Sharma et al., 1992). WTI expression is not completely down-regulated as WTI mRNA transcripts can be detected by PCR and the WTI protein is also present in the adult kidney (Mundlos et al., 1993).

The differentiation of nephrons is thought to occur in a similar way in all three kidneys. A similar pattern of expression is seen during nephron differentiation in the mesonephros and the metanephros in both man and mouse. Up-regulation of WTI occurs during tubule differentiation, declining as regression of the mesonephros takes place (Pritchard-Jones et al., 1990, Armstrong et al., 1992). The earliest nephrons in the pronephros are different in that there is no Bowman's capsule but filtration is through the coelomic epithelia into the coelom. These structures are known as an external glomeruli (Figure 1.5b). In these the pattern of WTI expression is not known.
Figure 1.4 Differentiation of the metanephros

a The ureteric bud within the metanephric mesenchyme divides to form the collecting ducts of the mature kidney and induces the differentiation of the mesenchyme into the nephrons from the centre of the kidney outwards. (Illustration from Gilbert 1991)

b Structure of the adult kidney, showing the division of the kidney into lobes and the cortex and the medulla. The glomeruli and proximal tubules are contained in the cortex and the collecting ducts and loops of Henle form the medulla. (Illustration from Kent 1987)
because of their very transient nature.

Vascularisation of the differentiating metanephric kidney is thought to occur mainly by invasion of cells from the surrounding capillaries (Saxen 1987) forming the glomeruli and the peritubular capillaries. But there is also by some angiogenesis within the kidney as demonstrated by capillary formation occurring in culture in the absence of the surrounding capillaries.

1.5.2 The molecular basis of kidney differentiation

The series of changes which occur within the cells as nephron development proceeds is only just becoming elucidated at the molecular level. The inductive signals from the nephric duct are thought to be permissive as opposed to instructive because other tissues can act as inducers of nephrogenesis demonstrating that the mesenchyme is predetermined (Sariola et al., 1989). Following induction by the ureteric bud the mesenchymal cells divide and this is followed by a change in surface markers. These changes are associated with alterations in the adhesion properties of the cells which are very important for the differentiation of the epithelial nephrons from mesenchyme. Thirty six hours after induction typically epithelial markers such as laminin A chain and \( \alpha \)Vomorulin (\( L-CAM \)) are detected. These are involved in the formation of a basement membrane and the formation of cell to cell contacts and the polarisation of the cells. The mesenchymal marker protein vimentin is lost at this time (Ekblom 1981).

The control of these events is still unclear. The expression of several transcription factors is up-regulated as the cells are starting to differentiate; \( N\text{-}myc \) (24 hours after induction) (Mugrauer and Ekblom 1991) and then \( PAX2, PAX8 \) and \( WTI \) (36 hours after induction) (Dressler et al., 1990, Plachov et al., 1990). Therefore the localisation and timing of expression of these genes points to a role in the control of the induction of nephrogenesis; specifically the mesenchymal to epithelial cell type transition taking place in the developing nephrons (Pritchard-Jones et al., 1990). The importance of one of these genes, \( PAX2 \), has been experimentally demonstrated, as antisense oligonucleotides in culture can block the cell type transition (Rothenpieler and Dressler 1993) and overexpression produces deformed kidneys in mouse, which resemble the condition congenital nephrotic syndrome (Dressler et al., 1993). It is thought that the \( PAX2 \) gene may be a target for \( WTI \) regulation because in Wilms' tumours \( PAX2 \) expression similarly to \( WTI \) expression is not down regulated as it is in normal kidney after tubular differentiation (Dressler and Douglas 1992). A fourth
Figure 1.5 Nephron development

a Mesenchyme is induced to condense by the nephric duct or the ureteric bud. The cells become polarised, form a basement membrane and take on the characteristics of epithelia. The condensate develops a lumen becoming comma and then S-shaped as the renal tubule forms. The end of the nephron nearest the collecting duct fuses to it while other end differentiates into the Bowman's capsule. Vascularisation occurs forming the glomeruli and the peritubular capillaries. (Illustration from Saxen 1987)
Figure 1.5 Nephron development (continued)

b In most mammalian nephrons the glomerulus is within the Bowman's capsule. Occasionally the nephron communicates to the coelom via a nephrostome. In the first pronephric tubules and lower vertebrates the glomerulus is in direct contact with the coelom (external glomerulus) with the coelomic epithelium equivalent to the podocytes of the Bowman's capsule. The nephric tubules also connect to the coelom. Therefore the formation of the Bowman's capsule is thought to have occurred by invagination of the glomerulus by the nephric tubule (internal glomerulus). (Illustration adapted from Kent 1987)
gene has also been implicated in this transition, *c-met* which is the ligand for hepatocyte scatter factor (Tsarfaly *et al.*, 1994).

1.5.3 Wilms' tumours

Wilms' tumours can be classified on the basis of histology. The most common type is triphasic, containing undifferentiated blastema, but also stroma, and elements which resemble the tubules and glomeruli. Heterologous tissues such as skeletal muscle, cartilage, adipose tissue and bone are also observed and are believed to be derived from the stroma. It is thought that after not taking the correct developmental pathway these undifferentiated cells can respond to inductive signals to which they are subsequently exposed to try to form nephrons or other differentiated tissue types. Tumours are also classified according to their association with 'islands' of abnormally persistent blastema known as nephrogenic rests. It has been proposed that these arise because the system of kidney differentiation depends on a series of inductions and is not determined by an inalterable 'blueprint' and is therefore susceptible to error (Mierau *et al.*, 1987). If the blastema either does not receive the inductive signal or fails to respond correctly to induction it can persist postnatally within the developed metanephric kidney. These are divided into either perilobar (PLNR) or intralobar (ILNR) types i.e. those found at the margins of the lobes of the metanephros or within the lobes (Beckwith *et al.*, 1990) (Figure 1.4b). ILNR associated tumours are often triphasic and contain heterologous tissues. PLNR associated tumours have a much more homogeneous histology, consisting mostly of blastema and epithelial elements. Since the metanephric blastema differentiates from the centre outwards, the peripheral location indicates that PLNR associated tumours have arisen later in nephrogenesis. At this time the cells maybe less pluripotent than in the initial wave of nephron formation resulting in the persistent blastema having fewer developmental options than in ILNR associated tumours. Nephrogenic rests have been reported in <1% of infant post-mortem (Bennington and Beckwith 1975). In contrast to the general population, rests are present in the kidneys of 40% of unilateral and 100% of bilateral Wilms' tumour patients (Bove and McAdams 1976, Beckwith *et al.*, 1990). The high incidence of WT and rest association and their similarities have suggested that tumours form from rests that have become neoplastic. In the case of ILNRs, these are often found at the edge of tumours indicating that the tumour may have arisen from a cell within the rest. Further evidence for this comes from two cases where the same WTI mutation was found in an associated rest as in the Wilms' tumour (Park *et al.*, 1993b). There is also a preferential association of rest and tumour type with different
associated congenital syndromes (Table 1.1). Both DDS and WAGR associate with the intralobar type (Mierau et al., 1987, Heppe et al., 1991) but perilobar types are more often to be found in BWS cases. This may indicate that there is genetically determined heterogeneity in the developmental abnormalities.

The expression of \textit{WTI} has been investigated in tumours and \textit{WTI} mRNA transcripts were found at high levels, but only in the analogous structures to those expressing \textit{WTI} in normal development. Tumours with a high blastemal content in general express high levels of \textit{WTI}. In more differentiated tumours \textit{WTI} becomes localised to the immature tubules and glomeruloid bodies (Pritchard-Jones and Fleming, 1991, Miwa et al., 1992a). Stromal rich tumours including stroma derived heterologous tissues (e.g. muscle and cartilage) were found to express very low or undetectable levels of \textit{WTI} (Pritchard-Jones and Fleming, 1991, Huang et al., 1990, Miwa et al., 1992a, Yeger et al., 1992). The origin of stromal tumours is not as clear as for other types. It is not known whether the precursor cells are derived from the uninduced or induced metanephric blastema. Pritchard-Jones and Fleming (1991) have suggested the second alternative on the basis that in order for a \textit{WTI} lesion to affect cell growth, the cells should have passed though the stage of \textit{WTI} expression. Mesenchymal cells which have not passed though the induced blastemal stage have not expressed \textit{WTI}.

In very rare cases tumours histologically identical to nephroblastoma occur outside the metanephric kidney and in the adult (Aterman 1989). Extra-renal nephroblastoma is mainly found in the retroperitoneal space and in the gonads, or their associated ducts. Such tumours were originally thought to be due to displaced metanephric tissue, but the tumours could possibly arise from mesonephric kidney remnants that have failed to regress or from the gonads themselves as they are very closely related to the kidneys (Pritchard-Jones and Hastie 1990). The expression of \textit{WTI} in these tumours has been investigated and it was found that in the endometrial derived tumours there is a persistence of high levels of \textit{WTI} expression similar to some Wilms' tumours indicating that \textit{WTI} may also be involved in their formation (Roberts et al., 1993).

1.6 \textit{WTI} and intermediate mesoderm development

The gonads develop in close juxtaposition to the mesonephroi in the intermediate mesoderm (Figure 1.3a&1.6a) and in addition it has been shown that for complete development of the gonad some interstitial cells must be contributed by the
Figure 1.6 Secondary sexual determination

a The indifferent gonad differentiates on the medial side of the mesonephros. The Mullerian duct differentiates in both sexes in the intermediate mesoderm parallel with the Wolffian (mesonephric) duct and both ducts join the cloaca.

b In the male the testis produces testosterone and AMH which induces the retention and differentiation of the Wolffian duct into the vas deferens and the mesonephric tubules into the epididymis and the regression of the Mullerian duct.

c In females the mesonephric tubules and ducts degenerate and the Mullerian duct form the oviducts and the uterus under the influence of oestrogen. (Illustrations from Gilbert 1991)
overlying mesonephros (Byskov 1986, Buehr et al., 1993). The non germ cell component of the gonad forms from a thickening of the coelomic epithelium in the intermediate mesoderm, differentiating into the interstitial component (including Leydig cells) and the epithelial component of the mature gonad involving a mesenchymal to epithelial cell type transition. From the first stages of development the gonad expresses high levels of WT1. On sexual differentiation this becomes localised to sex cords, in particular the Sertoli and granulosa cells of the adult gonad but not the germ cells (Pelletier et al., 1991c, Armstrong et al., 1992). WT1 expression is not as down regulated in the adult gonad as in the kidney as it remains detectable by northern blotting (Pelletier et al., 1991c). In the testis a smaller 2.5kb transcript was noted but not in other tissues (Pelletier et al., 1991c). Since then, the presence of a smaller transcript has been observed in other tissues and species (Sharma et al., 1992).

The mesothelium which covers all the internal organs also expresses WT1 during development and is derived from the coelomic epithelium. The kidney, gonads and mesothelium are related by their derivation from the coelomic epithelium, arising from the intermediate mesoderm and they all undergo a cell type transition, mesenchyme to epithelia, during development. Therefore they are intimately related by virtue of position, cell origin, differentiation pathway and the expression of WT1, indicating that many similar developmental decisions are taking place in these cells (Pritchard-Jones et al., 1990). The evolutionary origin of the glomerular podocytes is directly related to the mesothelium because in the primitive external glomeruli the filtration surface is not the podocytes but the coelomic epithelium. The Bowman's capsule seems to have been formed by the invagination of the glomeruli by the nephric tubules (Figure 1.5b). Interestingly the other organ to be derived from the intermediate mesoderm, the adrenal cortex, does not express WT1.

The genital ducts are also intimately associated with this region. In both sexes a second pair of ducts, the Mullerian ducts, forms from a longitudinal groove of the coelomic epithelium paralleling the mesonephric duct (Gilbert 1991). At this stage these ducts have not been observed to express WT1. Secondary sex determination which decides the fate of the Mullerian and Wolffian ducts (mesonephric ducts) is dependent on hormones produced by the developing testes and ovaries.

The developing testes produce testosterone from the Leydig cells and anti-Mullerian hormone (AMH) from the Sertoli cells. The former promotes the differentiation of the Wolffian ducts and the mesonephroi into the vas deferens and epididymis and the latter promotes the degeneration of the Mullerian ducts. (Figure
1.6b). Testosterone or its derivatives also affect the development of the external genitalia by acting on the urogenital sinus and swellings which then become the penis and scrotum (O'Rahilly 1977). After the initial development of the mesonephros none of these structures have been observed to express WTI (Pelletier et al., 1991c).

In the female in the absence of these two hormones, and in the presence of oestrogen produced by the ovaries, the Wolffian ducts and mesonephroi regress and the Mullerian ducts differentiate into the internal genitalia (Figure 1.6c). The caudal ends of the ducts open into the cloaca and the anterior ends remain open to form the ostia, which collects the mature oocytes. In the middle their fusion produces the uterus and contributes to the vagina (Jiresek 1977). Of these structures the ostia, the myometrium and endometrial stroma of the uterus but not the epithelial component of the endometrium have been observed to express WTI in the mouse and rat (Pelletier et al., 1991c, Zhou et al., 1993). During normal hormonal cycling in the female the expression of WTI was found to remain constant, but 6 days after fertilisation as the myometrial sma differentiated into the decidual cells, up-regulation of expression occurred, implicating WTI in differentiation of these cells.

1.7 WTI expression outside the intermediate mesoderm

Other cell populations with the same cell type transition have not been observed to express WTI by in situ hybridisation e.g. the somites, the endothelial lining of the capillaries, the corneal epithelium and the outer lining of the coelom derived from the lateral plate mesoderm (Armstrong et al., 1992). By northern blotting heart, lungs, and thymus (Buckler et al., 1991) express WTI but by in situ hybridisation this has been localised to their mesothelial covering (Armstrong et al., 1992, Park et al., 1993a).

At 11dpc the central region of the spinal cord in mouse expresses WTI and at day 15 this becomes specific to the ventral horn, this is also seen in human and rat. In the brain a small part of the medulla oblongata expresses, as does the area postrema in rats. Other non mesoderm derived structures, the tongue and the eye, express WTI when examined by RT-PCR (Armstrong et al., 1992, Sharma et al., 1992).

The first WTI clone was isolated from a preB-cell library (Call et al., 1990), so a possible role in haematopoiesis has been investigated. WTI expression has been shown to be limited to the splenic capsule and stroma, not the haematopoietic cells (Park et al., 1993a). Expression does occur in some immature leukaemic cells i.e. ALL, AML, CML but not in mature types (Call et al., 1990, Miwa et al., 1992b).
These regions of WTI expression outside the intermediate mesoderm and in structures not undergoing the mesenchyme epithelial transition indicate that WTI may have additional roles in other cell populations.

1.8 Involvement of WTI in genitourinary abnormalities

The association of Wilms' tumour with genitourinary abnormalities is particularly interesting because of the intimate association of kidney and genital systems and because both of them express high levels of WTI during development.

1.8.1 Mapping of genitourinary abnormalities to the WTI region

Mapping of constitutional WAGR deletions demonstrated that the locus involved in genitourinary abnormalities was in the same 345kb region that contained WTI. At the same time this clearly demonstrated the separate identity of the gene causing aniridia. It was proposed that genitourinary abnormalities could be due to a pleiotropic effect of WTI as opposed to the deletion of another gene, giving rise to this phenotype by haploinsufficiency (van Heyningen et al., 1990). The lack of one functional copy of WTI in the developing genitourinary system could perturb developmental events such as the production of the hormones from the gonads (Pritchard-Jones et al., 1990, Pelletier et al., 1991a) which control the development of the external genitalia therefore leading to these malformations (Behringer et al., 1990). This was further supported by two constitutional intragenic deletions in WTI being reported in two patients with Wilms' tumour, hypospadias and cryptorchidism but no aniridia (Pelletier et al., 1991a). As with WT, this cannot be a fully penetrant effect because one of these cases was familial but the father from whom the mutation was derived was unaffected. Several other cases of constitutional loss of one functional copy of WTI not resulting in genitourinary abnormalities have been reported (Table 1.2). The frequency of genitourinary malformations in XY WAGR is greater than for XX cases (Table 1.3). This could be consistent with the fact that the genitalia of females are not as dependent on hormone production from the developing gonad as males and so the perturbations caused by loss of one copy of WTI do not affect female genitalia development to such a great extent. This may be a reason why familial WT represents 1% of all cases (c.f. 35-45% for retinoblastoma (Knudson 1971)) because even where no obvious abnormalities are present fertility may be reduced.
1.8.2 Denys-Drash Syndrome

In addition to WAGR syndrome, Wilms' tumour is a component of Denys Drash syndrome which has much more severe genitourinary abnormalities in both XX and XY cases including ambiguous genitalia, pseudo-hermaphroditism, micropenis, cryptorchidism, hypospadias, and streak gonads (Denys et al., 1967, Drash et al., 1970). The kidneys also are affected and lose their function in childhood due to glomerular nephropathy (specifically mesangial sclerosis). Analysis of 10 DDS cases revealed constitutional point mutations in zinc fingers 2 and 3 of WTI, in these cases the mutations became homozygous in all the Wilms' tumours examined (Pelletier et al., 1991b). The majority of DDS patients examined have a specific set of mutations. These are missense mutations in zinc fingers 2 and 3 (Table 1.4 and references therein). These mutations affect the residues chelating the zinc ion or those which had been shown to interact with DNA in the co-crystal structure of EGR1 with its consensus sequence (Pavletich and Pabo 1991) (Figure 2). Approximately 60% of these are at R394 or D369 in ZF3. By analogy with the crystal structure of EGR1, the arginine residue probably binds a guanine base and the aspartate residue interacts with the arginine to stabilise the interaction with the DNA. Some of these mutations have been shown experimentally to destroy DNA binding to the EGR1 consensus sequence (Pelletier et al., 1991b). One interesting exception has been reported twice (Bruening et al., 1992, Konig et al., 1993) where a mutation in the splice donor site of exon 9 prevents the inclusion of the +KTS alternative splice. That this mutation in the heterozygous state can give rise to the DDS phenotype indicates that the ratio of the different isoforms of WTI must be of functional importance.

1.8.3 Dominant-negative hypothesis

The genitourinary abnormalities produced by these DDS mutations are more severe than those in WAGR cases and other constitutional deletion cases (Table 1.1). WAGR mutations normally lead to truncation of WTI or one WTI allele is completely deleted (Table 1.3) whereas DDS mutations are mostly missense mutations (Table 1.4). WAGR genitourinary abnormalities presumably are a result of WTI haploinsufficiency, while it is suggested that the DDS phenotype is due to a trans-dominant or dominant-negative effect (Herskowitz 1987, Su et al., 1993). A dominant effect would be due to a new function of the WTI mutant causing the phenotype, and a dominant-negative effect where the remaining normal WTI function is interfered with by the mutant (Pelletier et al., 1991b, Little et al., 1993). The
<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Tumour incidence</th>
<th>Karyotype</th>
<th>Mutation in WT1 gene</th>
<th>Change in WT1 protein</th>
<th>Status of mutation in tumour</th>
<th>First hit</th>
<th>Second hit</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP, C</td>
<td>bilateral sporadic</td>
<td>46XY</td>
<td>17bp deletion in exon 4</td>
<td>truncation in exon 4</td>
<td>1) homozygous 2) homozygous</td>
<td>WT1 truncation</td>
<td>LOH</td>
<td>Pelletier et al., 1991a</td>
</tr>
<tr>
<td>HP, C</td>
<td>unilateral Familial</td>
<td>46XY</td>
<td>1bp deletion in exon 6</td>
<td>truncation in exon 7</td>
<td>homozygous</td>
<td>WT1 truncation</td>
<td>LOH</td>
<td>Pelletier et al., 1991a</td>
</tr>
<tr>
<td>HP, C, DK, AN</td>
<td>sporadic</td>
<td>46XY del(11)p13p14.1</td>
<td>deletion of exons 2-3 in mRNA</td>
<td>truncation after exon 4</td>
<td>hemizygous</td>
<td>large constitutional deletion</td>
<td>WT1 truncation</td>
<td>Brown et al., 1992</td>
</tr>
<tr>
<td>GU, R, AN</td>
<td>unilateral sporadic</td>
<td>N/D del(11)p12p13</td>
<td>10bp insertion in exon 7</td>
<td>truncation in exon 7</td>
<td>hemizygous</td>
<td>large constitutional deletion</td>
<td>WT1 truncation</td>
<td>Baird et al., 1992a</td>
</tr>
<tr>
<td>GU, R, AN</td>
<td>unilateral sporadic</td>
<td>N/D del(11)p13</td>
<td>chain termination mutation in exon 8</td>
<td>362R-STOP in ZF2</td>
<td>hemizygous</td>
<td>large constitutional deletion</td>
<td>WT1 truncation</td>
<td>Baird et al., 1992a</td>
</tr>
<tr>
<td>HP, C, AN</td>
<td>unilateral sporadic</td>
<td>46XY del(11)p13p14.1</td>
<td>1bp deletion in exon 7</td>
<td>truncation in exon 7</td>
<td>hemizygous</td>
<td>large constitutional deletion</td>
<td>WT1 truncation</td>
<td>Gessler et al., 1993</td>
</tr>
<tr>
<td>C, AN, R</td>
<td>unilateral sporadic</td>
<td>46XY del(11)p13</td>
<td>missense mutation in exon 3</td>
<td>201G-D</td>
<td>hemizygous</td>
<td>large constitutional deletion</td>
<td>WT1 truncation</td>
<td>Park et al., 1993c</td>
</tr>
<tr>
<td>AGU, AN, R</td>
<td>unilateral sporadic</td>
<td>46XX del(11)p12p14.3</td>
<td>14bp duplication at exon/intron7 splice donor site</td>
<td>uncharacterised -may be exon skipping leading to truncation</td>
<td>hemizygous</td>
<td>large constitutional deletion</td>
<td>WT1 mutation</td>
<td>Santos et al., 1993</td>
</tr>
<tr>
<td>BWS</td>
<td>unilateral sporadic</td>
<td>46XX</td>
<td>missense mutation in exon 2</td>
<td>154F-S</td>
<td>heterozygous</td>
<td>genetic event at WT2</td>
<td>WT1 mutation</td>
<td>Park et al., 1993b</td>
</tr>
</tbody>
</table>

AGU, absence of gonads and uterine cavity; AN, aniridia; BWS, Beckwith Wiedemann Syndrome; C, cryptorchidism; DK, duplex kidney; GU, unspecified genitourinary abnormalities; HP, hypospadias; R, mental retardation.
### Table 1.4 WTI mutations associated with Denys-Drash syndrome

Summary of constitutional WTI mutations found in patients with DDS.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Karyotype</th>
<th>WTI mutation in gene</th>
<th>Change in WTI protein</th>
<th>Incidence of Wilms' tumour</th>
<th>Status of mutation in tumour</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F, SG</td>
<td>46XY</td>
<td>missense mutation in ZF2</td>
<td>366R-H</td>
<td>none</td>
<td>gonadalblastoma</td>
<td>Pelletier et al., 1991b</td>
</tr>
<tr>
<td>AM or F, SG</td>
<td>Two, 46XY Two, 46XX</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>unilateral</td>
<td>Homozygous in two cases examined</td>
<td>Pelletier et al., 1991b</td>
</tr>
<tr>
<td>AM</td>
<td>N/D</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>bilateral</td>
<td>homozygous</td>
<td>Pelletier et al., 1991b</td>
</tr>
<tr>
<td>F</td>
<td>46XX</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>bilateral</td>
<td>N/D</td>
<td>Pelletier et al., 1991b</td>
</tr>
<tr>
<td>AM</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>396D-G</td>
<td>none</td>
<td>-</td>
<td>Pelletier et al., 1991b</td>
</tr>
<tr>
<td>F</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>396D-N</td>
<td>none</td>
<td>-</td>
<td>Pelletier et al., 1991b</td>
</tr>
<tr>
<td>F, SG</td>
<td>46XY</td>
<td>splice acceptor site mutation in intron 9</td>
<td>KTS alternative splice not included</td>
<td>none</td>
<td>-</td>
<td>Bruening et al., 1992</td>
</tr>
<tr>
<td>F</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>394R-P</td>
<td>unilateral</td>
<td>homozygous</td>
<td>Bruening et al., 1992</td>
</tr>
<tr>
<td>AM, SG</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>unilateral</td>
<td>N/D</td>
<td>Bruening et al., 1992</td>
</tr>
<tr>
<td>CL</td>
<td>46XX</td>
<td>missense mutation in ZF1</td>
<td>330C-Y</td>
<td>none</td>
<td>-</td>
<td>Bruening et al., 1992</td>
</tr>
<tr>
<td>F</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>none</td>
<td>-</td>
<td>Little et al., 1993, Baird et al., 1992b</td>
</tr>
<tr>
<td>F</td>
<td>46XX</td>
<td>missense mutation in ZF3</td>
<td>396D-N</td>
<td>bilateral</td>
<td>both homozygous</td>
<td>Little et al., 1993</td>
</tr>
<tr>
<td>F</td>
<td>N/D</td>
<td>missense mutation in ZF3</td>
<td>360C-G</td>
<td>unilateral</td>
<td>heterozygous</td>
<td>Little et al., 1993</td>
</tr>
<tr>
<td>MP, C, HK</td>
<td>46XY</td>
<td>chain termination mutation in ZF2</td>
<td>truncation in exon 8</td>
<td>bilateral</td>
<td>both homozygous</td>
<td>Little et al., 1993</td>
</tr>
<tr>
<td>HP</td>
<td>46XY</td>
<td>missense mutation in ZF2</td>
<td>373H-Q</td>
<td>none</td>
<td>-</td>
<td>Little et al., 1993</td>
</tr>
<tr>
<td>GU</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>unilateral</td>
<td>N/D</td>
<td>Baird et al., 1992b</td>
</tr>
<tr>
<td>GU</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>unilateral</td>
<td>homozygous</td>
<td>Baird et al., 1992b</td>
</tr>
<tr>
<td>NM</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>396D-N</td>
<td>unilateral</td>
<td>homozygous</td>
<td>Baird et al., 1992b</td>
</tr>
<tr>
<td>AM</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>366R-H</td>
<td>unilateral</td>
<td>N/D</td>
<td>Baird et al., 1992b</td>
</tr>
<tr>
<td>AM</td>
<td>46XY</td>
<td>insertion 1bp in exon 6</td>
<td>truncation in exon 6</td>
<td>bilateral</td>
<td>homozygous</td>
<td>Baird et al., 1992b</td>
</tr>
<tr>
<td>AM</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>none</td>
<td>-</td>
<td>Coppes et al., 1992b</td>
</tr>
<tr>
<td>F</td>
<td>46XX</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>bilateral</td>
<td>N/D</td>
<td>Coppes et al., 1992b</td>
</tr>
<tr>
<td>F, SG</td>
<td>46XY</td>
<td>missense mutation in ZF2</td>
<td>377H-Y</td>
<td>none</td>
<td>-</td>
<td>Coppes et al., 1992b</td>
</tr>
<tr>
<td>F, SG</td>
<td>46XY</td>
<td>splice acceptor site mutation in intron 9</td>
<td>KTS alternative splice not included</td>
<td>none</td>
<td>-</td>
<td>Konig et al., 1993</td>
</tr>
<tr>
<td>F, SG</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>unilateral</td>
<td>N/D</td>
<td>Sakai et al., 1993</td>
</tr>
<tr>
<td>F</td>
<td>46XY</td>
<td>missense mutation in ZF3</td>
<td>394R-W</td>
<td>unilateral</td>
<td>N/D</td>
<td>Sakai et al., 1993</td>
</tr>
<tr>
<td>F</td>
<td>46XX</td>
<td>missense mutation in ZF2</td>
<td>355C-Y</td>
<td>none</td>
<td>-</td>
<td>Sakai et al., 1993</td>
</tr>
<tr>
<td>F</td>
<td>46XX</td>
<td>missense mutation in ZF3</td>
<td>401H-Y</td>
<td>unilateral</td>
<td>N/D</td>
<td>Baird and Cowell 1993</td>
</tr>
<tr>
<td>HP, C</td>
<td>46XY</td>
<td>insertion 1bp in exon 9</td>
<td>truncation in ZF3</td>
<td>none</td>
<td>-</td>
<td>Ogawa et al., 1993b</td>
</tr>
<tr>
<td>AM, HP, C</td>
<td>46XY</td>
<td>splice donor site mutation in exon 6</td>
<td>Exon skipping deletes exon 6 truncation in exon 7</td>
<td>unilateral</td>
<td>homozygous</td>
<td>Schneider et al., 1993</td>
</tr>
<tr>
<td>F, CL</td>
<td>46XY</td>
<td>missense in ZF2</td>
<td>360C-Y</td>
<td>none</td>
<td>-</td>
<td>Clarkson et al., 1993</td>
</tr>
<tr>
<td>HP, C</td>
<td>46XY</td>
<td>missense in ZF2</td>
<td>377H-R</td>
<td>unilateral</td>
<td>N/D</td>
<td>Nordenskjold et al., 1994</td>
</tr>
<tr>
<td>HP, C</td>
<td>46XY</td>
<td>missense in ZF3</td>
<td>396D-N</td>
<td>unilateral</td>
<td>homozygous</td>
<td>Nordenskjold et al., 1994</td>
</tr>
<tr>
<td>F, GD</td>
<td>46XX</td>
<td>missense in ZF3</td>
<td>394R-W</td>
<td>none</td>
<td>-</td>
<td>Nordenskjold et al., 1994</td>
</tr>
<tr>
<td>F, GD</td>
<td>46XX</td>
<td>missense in ZF3</td>
<td>396D-N</td>
<td>unilateral</td>
<td>homozygous</td>
<td>Nordenskjold et al., 1994</td>
</tr>
</tbody>
</table>

AM, ambiguous genitalia; C, cryptorchidism; GD gonadal dysgenesis; GU, unspecified genitourinary malformations; HK, horseshoe kidney; CL, clitoromegaly; HP, hypospadias; MP, micropenis; NF, female external genitalia; NM, normal male genitalia; SG, streak gonads.

JGCT, juvenile granulosa cell tumour
evidence for the DDS phenotype being caused by a dominant-negative type of mutation, as opposed to a purely dominant one, comes from the observation that truncation of the WTI zinc fingers (which leaves the transregulatory and dimerisation domains intact) as well as the zinc finger missense mutations can lead to more severe genitourinary defects (Baird et al., 1992b, Little et al., 1993, Schneider et al., 1993). If WTI functions only as an intact dimer, then such mutations might be expected to reduce the functional dimer concentration to 25%. Initially it was thought that DDS and WAGR phenotypes were due to different types of WTI mutants. This is not exclusively the case as intermediate phenotypes are now being observed e.g. an XX WAGR case with an 11p12-13 deletion but with nephropathy which is normally characteristic of DDS (Baird et al., 1992b), and other cases lacking the intersex disorders but still showing the diagnostically characteristic glomerular nephropathy (Habib et al., 1985). The same deletion of 11p13 was inherited as an unbalanced translocation in two cousins, the XX case had only aniridia but the XY case had DDS like genitourinary abnormalities (Henry et al., 1993b). Therefore it may be that WAGR and the DDS represent different parts of a spectrum of genitourinary abnormalities dependent on the type of mutation, genetic background and a higher susceptibility in XY cases. Like WAGR, DDS may be incompletely penetrant. Two phenotypically normal cases where Drash-type WTI mutations are present constitutionally have been observed, a girl with Wilms' tumour (Akasaka et al., 1993) and the father of a DDS son (Coppes et al., 1992b). In all the other cases where inheritance was looked for parental mutations were not found; phenotypically normal carriers of Drash-type mutations are probably very rare. It is now becoming clear that different mutations in a single gene can give a range of effects on the developmental pathways. This has been found for PAX6 mutations where a range of eye phenotypes are obtained (Hanson et al., 1994) and for the ret oncogene (van Heyningen 1994).

1.9 Involvement in other developmental abnormalities

Since WTI is expressed at sites other than the kidney, the possible role of WTI in tumorigenesis of gonadal tumours and mesotheliomas has been investigated. One case has been reported of a cystic peritoneal mesothelioma in which a homozygous point mutation was found. It is debatable whether this is a tumour or a developmental abnormality and as yet no other mesotheliomas have been detected with mutations in WTI. The mutation found in exon 6 (codon 273) resulted in a serine to glycine substitution and has been shown to change the transrepressor activity of WTI into
transactivation activity in *in vitro* systems (Park *et al.*, 1993a). Mutations, such as this, that do not delete or change the sequence of the zinc fingers are so far very rare.

Gonadoblastoma has been observed in a few cases of DDS in which *WT1* point mutations have been found (Pelletier *et al.*, 1991b, Table 1.4). *WT1* mutations are probably not a major cause of these tumours as no mutations have been seen in Frasier's syndrome (nephrotic syndrome and gonadoblastoma) (Poulat *et al.*, 1993), ovarian tumours (Bruening *et al.*, 1993), sex cord-stroma tumours (Coppes *et al.*, 1993b) or other urogenital tumours (Quek *et al.*, 1993). In the ovarian tumours an up-regulation of expression was observed, similar to some Wilms' tumours which have no detectable *WT1* mutations (Bruening *et al.*, 1993).

### 1.10 The *Wt1* knockout mouse

The ultimate demonstration of the importance of a gene in a developmental event is the failure of this event to occur when the gene is homozygously deleted. Homozygous deletion of the 5' end of the *Wt1* gene leads to the failure of kidney and gonad development and embryonic lethality in mice. Mesonephric differentiation occurs, but 2-3 fold fewer tubules form. Formation of the uninduced metanephric blastema is also observed but no ureteric bud develops so that induction does not happen and the blastema undergoes apoptosis. The initial thickening of the coelomic epithelium at the onset of gonad differentiation appears but again development does not proceed beyond this point. The embryonic lethality is probably not due to the above, but to heart, lung, and diaphragm abnormalities which are also observed, and may be ascribed to a failure of the mesothelium which covers these organs to develop fully (Kreidberg *et al.*, 1993). This indicates that *Wt1* is absolutely required for murine kidney and gonad development and also suggests that the expression in the coelomic epithelium and later in the mesothelium also plays a key role in the normal development of heart, lungs, and diaphragm. The heterozygously deleted mice did not develop WT (Kreidberg *et al.*, 1993), this indicates that there are some distinct differences in the development of mouse and human kidneys.

### 1.11 Is *WT1* the only gene responsible for predisposition to Wilms' tumour?

#### 1.11.1 Lack of mutations in *WT1*

Although it is clear that the homozygous loss of *WT1* function is the mechanism of tumorigenesis in some Wilms' tumours (summarised in Tables 1.2-1.4), it cannot be
the only mechanism. Only 1-2% of WT patients have visible cytogenetic deletions in 11p13 and only 30% show LOH (Little et al., 1991c). The WT1 gene was found to be unaltered in the genome as observed by Southern blotting (Cowell et al., 1991, Royer-Pokora et al., 1991), and WT1 expression was observed in many Wilms' tumours (Pritchard-Jones et al., 1990). Therefore it was expected that the gene would contain small changes resulting in a loss of function of the protein. Many Wilms' tumours have now been examined. Analysis of the literature shows that in only about 10% is there a loss of the functional gene, corresponding to Knudson's two hit model, with many tumours having no changes in WT1 in either allele (Baird et al., 1992a&b). This may be due to mutations outside of the transcribed or coding regions which may not have been analysed, affecting WT1 transcription or translation, but this is unlikely for so many cases. The one group that does seem to follow Knudson's model are the DDS cases, except in two cases (Little et al., 1993, K. Williamson personal communication) where the constitutional mutation in WT1 becomes homozygous in the Wilms' tumour (Table 1.4).

1.11.2 Dominant WT1 mutations and trans-splicing affects

The heterozygous loss of zinc finger 3 in the tumour AR has been further investigated to determine if this may have a dominant tumorigenic effect over the remaining wild type WT1 allele (Haber et al., 1990). The deleted WT1, but not wild-type WT1, can co-operate with adenovirus E1A protein to produce a transformed phenotype (Haber et al., 1992). This demonstrates that mutant WT1 can act in a dominant way, but not that it is responsible for the development of the Wilms' tumour. Another mechanism is now becoming apparent with the isolation of some WT1 transcripts lacking exon 2 from Wilms' tumours but not from normal kidney cells. No mutations have been detected in the genomic DNA indicating WT1 is affected by a trans-splicing mechanism (Haber et al., 1993).

1.11.3 A possible role for imprinting

It has been observed that the paternal allele is preferentially retained in Wilms' tumours, suggesting that tumorigenesis is favoured by the retention of this allele. It was proposed that this could be due to imprinting of WT1 (Schroeder et al., 1987, Williams et al., 1989, Pal et al., 1990). The expression of maternal and paternal alleles in kidney and tumour tissue has been shown, disproving this theory (Haber et al., 1990, Little et al., 1992b). Paternal alleles may be more frequently selected in
tumours because they are the ones which contain the mutation and therefore give the growth advantage. This is because mutations happen at a higher frequency during male meiosis and so will be transmitted through the paternal germline (Charlesworth 1993). Retention of the paternal allele could also be due the presence of a linked imprinted transforming gene which can synergise with \textit{WTI} to produce a transformed phenotype (Wilkins 1988).

1.11.4 A second Wilms' tumour locus on 11p

From loss of heterozygosity studies a second locus on 11p, 11p15.5, has been implicated in the predisposition to Wilms' tumour (Mannens \textit{et al.}, 1988, Henry \textit{et al.}, 1989, Reeve \textit{et al.}, 1989, Wadey \textit{et al.}, 1990). The locus for Beckwith-Wiedemann syndrome has now been mapped to 11p15.5 (Waziri \textit{et al.}, 1983, Koufos \textit{et al.}, 1989). BWS patients are susceptible to several tumours (Table 1.1), although more than 50\% are Wilms' tumours, indicating that there maybe a second tumour predisposing locus, \textit{WT2}.

\textit{IGF2} has been proposed as a candidate gene for BWS (Little \textit{et al.}, 1991b). It is located at 11p15.5 and has been shown to be imprinted in most tissues where it is expressed including the foetal kidney, lung, brain, muscle, liver, adult kidney and placenta (Kalscheuer \textit{et al.}, 1993, Ohlsson \textit{et al.}, 1993) with only the paternal allele being expressed. There is evidence for the overgrowth observed in BWS arising from an 'overdose' of \textit{IGF2}, which is a growth hormone, either owing to deregulation of imprinting (Weksberg \textit{et al.}, 1993), or paternal disomy (Henry \textit{et al.}, 1991, Henry \textit{et al.}, 1993a). This is consistent with the phenotype observed in mouse as overexpression of \textit{Igf2} gives normally proportioned but differently sized mice (DeChiara \textit{et al.}, 1991). That an 'overdose' of \textit{IGF2} is also involved in Wilms' tumour formation is suggested by the observation of deregulation of \textit{IGF2} imprinting (Ogawa \textit{et al.}, 1993a\&c) and LOH (Chao \textit{et al.}, 1993) in Wilms' tumours. There may have to be further genetic changes than simple LOH to obtain a phenotype, as 11p homozygous tissues have also been found to develop with a normal phenotype (Chao \textit{et al.}, 1993).

The loss of heterozygosity at 11p15.5 in Wilms' tumours as opposed to 11p13 suggests that changes at 11p15.5 (\textit{WT2}) can co-operate with loss of one copy of \textit{WTI} in tumorigenesis. Recently a heterozygous \textit{WTI} mutation was detected in a perilobar Wilms' tumour from a patient with BWS stigmata (Park \textit{et al.}, 1993b). This suggests that the first hit occurred at the \textit{WT2} locus and the second hit was in \textit{WTI}. This gives
weight to the suggestion that the *WT1* and *WT2* loci can co-operate in tumour formation.

1.11.5 Not all Wilms' tumour predisposition loci map to chromosome 11

Non random allele loss and frequent cytogenetic rearrangements have been observed on the long arm of chromosome 16 as well as on 11p, and so 16q has been proposed as the location of *WT3* (Coppes et al., 1992a, Maw et al., 1992). So far the most likely candidate gene is the epithelial protein, vomorulin.

Familial cases of Wilms' tumour are rare, although there are now cases of hereditary *WT1* mutations (Pelletier et al., 1991a), WAGR cases (Fantes et al., 1992, Henry et al., 1993b) and even one DDS mutation (Coppes et al., 1992b). In several families predisposition to WT is found not to segregate with 11p13 or 11p15 markers (Grundy et al., 1988, Huff et al., 1988, Schwartz et al., 1991) and in one study not with chromosome 16 either (Huff et al., 1992).

1.11.6 Progression of Wilms' tumours

Secondary mutational events are probably involved in the progression of Wilms' tumour once *WT1* or other Wilms' tumour predisposition loci have been affected. An obvious candidate is p53 which is affected in many human tumours and may interact with *WT1* (Maheswaran et al., 1993). p53 has been found to be highly overexpressed in Wilms' tumour, particularly in the epithelial elements (Lemoine et al., 1992). In a tumour which was without a *WT1* mutation, a mutation in intron 1 of p53 was found (Velasco et al., 1993). p53 mutations have now been specifically linked to the rare anaplastic type of Wilms' tumours, which are associated with poor prognosis (Bardeesy et al., 1994). Chromosomal abnormalities and allele loss have been found in tumours on 1, 4p, 6, 7, 8q, 12, 14q, 16, 17, and 18 (Wang-Wuu et al., 1990, Maw et al., 1992).

1.12 Evolutionary comparison of the vertebrate genitourinary system

Comparison of the development of the genitourinary systems within the vertebrates can be used to explore the relationships between different organs by examining the way in which they have evolved. The theme that runs through kidney and gonad development is their intimate association with each other and with the
coelomic epithelium. This close evolutionary relationship may explain the expression of WT1 in all three structures.

1.12.1 The vertebrate kidney

The kidney or holonephros is a characteristic feature of vertebrates. Holonephros is a term which includes all the nephrogenic tissue (Saxen 1987), and also refers to the archetypal vertebrate kidney (Kent 1987). The nephron in its various forms is found in all vertebrates, and nephron-like imprints have been found in fossils of the ostracoderms, the oldest known vertebrates (Torrey 1965), but not in any other sub-phyla of Chordata. The nephrons seem to have developed from specialisation of mesodermal ducts, known as coelomoducts, which often lead from the coelom. Coelomoducts of mesodermal origin do exist outside of Chordata, but do not resemble vertebrate kidneys (Barnes et al., 1993). In all vertebrates the differentiation of the intermediate mesoderm occurs as a wave of differentiation followed by degradation. The final structures formed depend on the extent of these two events (Figure 1.8).

1.12.2 Amniote kidney development

The amniotes (mammals, birds and reptiles) all develop three successive kidneys, the first two being lost during foetal life or early in adulthood (Figure 1.7&1.8). The pronephros is not functional, and whether it actually exists in mammals has been debated (Torrey 1965). This partly arises from the difficulty in defining boundaries between the kidneys, in what is a continuous wave of differentiation (Saxen 1987). The origin of the pronephric duct is not well understood in mammals. In the chick a distinct duct rudiment differentiates from the lateral plate mesoderm before any tubules are seen (P. Lear personal communication), as opposed to being derived from the fusion of the first pronephric tubules as has been proposed for amphibians (Torrey 1965). The pronephros is less transitory in reptiles although it is thought not to be functional (Fox 1977). The mesonephros has been shown to be the functional foetal kidney in the reptilian lineages (Romanoff 1960). Its degeneration occurs later in the reptilian lineages so that in some species of snakes it is present until after the first moult. This is in contrast to mammals where the mesonephros is vestigial and has disappeared well before birth.
The tree shows the separation of the fish, then amphibian, then reptilian lineages from the mammalian lineage. The separation of the bird lineage from the reptilian lineage happened after the separation from the mammalian lineage so that birds and some reptiles belong to the same clade. (Tree adapted from Powers 1991)
Figure 1.8 The relationship between kidney development and evolution

Vertebrate kidneys are all formed as a result of a wave of differentiation of the urogenital ridge of the intermediate mesoderm followed by a wave of degeneration. In fish and amphibians this can result in the formation of mesonephroi of different extents. In the reptiles, birds and mammals the mesonephros is replaced by the metanephros. (Illustration adapted from Kent 1987)
1.12.3 Anamniote kidney development

In anamniotes (amphibians and fish) the mesonephros is the adult kidney (sometimes called the opisthonephros) as no metanephros differentiates (Figure 1.8). The pronephros is the functional foetal kidney and in some teleosts it continues to function in the adult. Many more of the nephrons, especially the anterior ones, are found to have the more ancient external glomeruli (Figure 1.5b). The capacity for nephron development may not be limited to embryonic life in lower vertebrates (Fox 1977).

1.12.4 The metanephros

The feature that appears to be the evolutionary novelty that causes the metanephros to form is the ureteric bud. The region of nephrogenic tissue that forms the metanephros in amniotes, is capable of forming opisthonephric tubules when induced by the mesonephric duct in the lamprey, shark and urodele. The mesonephros in mammals contains 30-50 nephrons while the metanephric tubules are present in their millions. This may be a consequence of the mesonephric function becoming of lesser importance, but even in the opisthonephros only a few hundred nephrons are present (Kent 1987). The greater capacity of the metanephros could have been selected for by a greater requirement for fluid conservation with the colonisation of terrestrial habitats. Selection for greater fluid conservation within mammals by the metanephros is exemplified by the almost complete water conservation achieved by very long loops of Henle in desert species such as the kangaroo rat. A greater capacity for excretion may also have been involved in the evolution of homeothermy and the associated increase in metabolic rate and production of toxic wastes.

1.12.5 The gonads and genital ducts

In all vertebrates the gonads develop in very close proximity to the mesonephroi and to the coelom. The mechanism by which gonadal sex is determined is very variable. In fish and amphibia there is the capacity for sex reversal in adult life. In reptiles the mechanism of sex determination can be temperature dependent but can also be influenced genetically. In birds and mammals it is determined solely genetically (Gilbert 1991).
Figure 1.9 The interrelationship between the gonadal and urinary systems

In the earliest gnathostomes the kidney ducts became utilised by the gonads for transport of the sperm to give the basic genitourinary system. In the most primitive of fish this organisation has remained. In other groups of vertebrates sperm and urine transport have become separated by the production of a second duct, either to carry sperm (fish) or urine (amniotes). (Illustration adapted from Kent 1987)
In the most ancient vertebrates, the agnatha, there are no genital ducts and the gametes are discharged into the coelom and exit via abdominal genital pores. In the most primitive gnathomes the bridge between the testes and kidneys has been made and the urinary duct function has been usurped by the gonads. This situation is still found in some fish e.g. the sturgeon. In most other systems there has been a separation of sperm and urine transport by the formation of a second set of ducts. In the fish and urodeles these carry sperm and are formed as outgrowths from the testes and in amniotes the urine is transported by the ureter, formed from the ureteric bud (Romer and Parsons 1977) (Figure 1.8). In females a second set of ducts is formed in all species from the gnathomes onwards. In amphibians and elasmobranchii the relationship of the female genitalia to the urinary ducts is closer than in higher vertebrates, as the ostia are formed from the pronephric tubules and the Mullerian ducts from the longitudinal splitting of the pronephric duct, as opposed to a separate origin in the intermediate mesoderm. These ducts become specialised, forming shell glands, ovisacs, and uteri.

1.13 Evolutionary comparison of Wilms'-like tumours

1.13.1 Mammals

Although there is a high degree of similarity of the mouse \( \text{Wt1} \) gene to human \( \text{WT1} \) at both the predicted protein sequence and expression levels, very few nephroblastomas have been reported in the mouse (Liebelt et al., 1989). This is surprising considering the number of mice in laboratories. This observation is supported by the fact that the heterozygous SeyDeY (small eye) mice, which are deleted for the region corresponding to the human WAGR region (Glaser et al., 1990), have an eye phenotype similar to aniridia (Jordan et al., 1992), but there is no evidence of the genitourinary abnormalities or Wilms' tumours. In addition mice heterozygous for the \( \text{Wt1} \) deletion do not develop nephroblastomas (Kreidberg et al., 1993). This is also the case in retinoblastoma, where the mechanism of tumorigenesis differs between mouse and human, as retinoblastomas are not observed in mice heterozygous for a non-functional retinoblastoma gene (Jacks et al., 1992).

Reports of nephroblastomas in other animals tend to be rare and there is often the problem of ascertaining whether it is a true nephroblastoma and so comparable with Wilms' tumour or is derived from a different type of precursor cell (Hard 1984). The rat, has an almost identical \( \text{Wt1} \) gene to the mouse but has been shown to develop nephroblastomas with the characteristic blastemal cells and tubule and glomerular like
structures and certain strains have a higher level of predisposition (Hard and Noble 1981). Stromal and heterologous elements are also observed in mesenchymal tumours. These are not true nephroblastomas but have a histology similar to some human Wilms' tumours (Hard 1984). Both these tumours can be induced by exposure to nitroso compounds (Jasmin and Riopelle 1970). Wilms' like tumours have also been found in a number of eutherian mammals; rabbits, hamsters, pigs, cows, dogs, horses, sea-lion, elk, cats, sheep, goats, macaque and red monkey, and in a metatherian mammal; the opossum. In some species no age dependency for tumour development is observed (dogs), in others it is found at approximately the same time as humans (pigs), in others prenatally (cows) (Hard 1984). WTI expression has been observed in pigs and shows a similar down regulation in kidney expression in early life. In addition contra-lateral nephrectomy induced expression in adult pigs may be as a result of compensatory renal growth (Kushner et al., 1992). Nitroso compounds have been found to induce tumours in many of these species, the window for exposure corresponding to the presence of metanephric blastema (Jurgelski et al., 1976).

1.13.2 Birds

In fowl spontaneous nephroblastoma is one of the most common neoplasms, occurring mostly in young birds. In the laboratory nephroblastomas can be induced by infection with viruses, mainly avian myeloblastosis virus (AMV) (Heine et al., 1962). Although virus induced tumours and spontaneous ones are histologically identical no virus has been shown as yet to be associated with spontaneous ones (Heine et al., 1962, Ishiguro et al., 1962). The sites of viral integration have been examined and any changes occurring in RNA expression investigated. The gene that has been implicated is not WTI but novC (Joliot et al., 1992). In one case a truncated transcript was produced as a result of viral integration, in others expression was seen to be up-regulated in the resulting nephroblastomas. The predicted protein belongs to a cysteine rich secreted family of immediate growth response genes which show similarity to insulin-like growth factor 1 binding protein (IGF1BP) (Joliot et al., 1992). The truncated form of the gene has been shown to transform cells in culture but the full length transcript repressed growth. The human homolog NOVH has been isolated and maps to 8q24.1 proximal to c-myc. This is a region found to be affected in Wilms' tumours with above random frequency (Martinerie et al., 1992). NOVH may well have a role in human Wilms' tumour progression but this has yet to be demonstrated.
The nearest relatives to the aves are the crocodilians but there are no reports of nephroblastoma in the reptilian class. While this may be a true reflection of the situation, it is more likely that it is a rare tumour and has not yet been observed (Elkan 1963).

1.13.3 Anamniotes

Nephroblastomas do occur in classes only possessing the pronephros and mesonephros. In *Xenopus laevis* tumours consisting of undifferentiated blastema and tubules have been observed (Elkan 1963), and in the fire bellied newt (*Cynops pyrrhogaster*) primitive glomerular differentiation was seen (Zwart et al., 1970). Most of the nephroblastomas in fish have been observed in the *Salmonidae*. In the rainbow trout (*Salmo gairdneri*) tumour induction is possible by exposure of embryos to nitroso compounds (Hendricks et al., 1980). Recently there has been a report of characteristic Wilms'-type tumours in the Japanese eating eel (*Anguilla japonica*) containing blastemal cells, renal tubules and striated muscle (Masahito et al., 1992).

1.14 Evolutionary comparison as a method for defining gene function

1.14.1 Structural analysis

The rate of change of amino acid sequence can vary by as much as a thousand fold; ranging from the fast evolving fibrinopeptides to the virtually static histone H4 (Nei 1987). Conserved regions between the same gene in different species are thought to arise from functional and structural constraints on the gene and gene product. Originally it was thought that mutations would become fixed or removed from the population owing to their advantageous or disadvantageous phenotype. This would be consistent with Darwin's theory of natural selection. Kimura's neutral theory of molecular evolution (Kimura 1968, Kimura 1983) was developed in order to explain the way in which evolution occurs at the molecular level. This states that most mutations are neutral or nearly neutral and these become fixed in the population due to factors such as genetic drift and that Darwinian selection will only act on a subset of mutations. Some of the best evidence in favour of this is the evolution of third bases of codons in protein coding sequences (Kimura 1977). Because of the degeneracy of the amino acid code many nucleotide substitutions will be neutral or almost so (RNA structure or tRNA abundance can affect this), but these have been shown to be some of the most prevalent sites of nucleotide substitution. Even more rapidly evolving are regions of DNA with no known function e.g. pseudogenes (Li et
Therefore it is postulated that, assuming a gene retains a similar function, those regions that have been conserved through evolution are important for its structure and function.

The number of known gene sequences is increasing very rapidly, and so the opportunities for comparison of orthologs (genes related by speciation) and paralogs (genes related by duplication) also increases. In particular the number of transcription factors studied has expanded. Comparative sequence analysis has identified several important amino acid sequence motifs e.g. the heptad repeat of leucines in the leucine zipper (Landshultz et al., 1988), the POU domain (Strum and Herr 1988) and the REL homology domain (Kerr et al., 1993). Also comparison of newly isolated genes containing known motifs may give clues to the characteristics and function of the gene product. In this way many genes are now being grouped into families and superfamilies e.g. AP1 family (containing the fos and jun transcription factors) (reviewed by Lamb and McKnight 1991) and the immunoglobulin superfamily. In addition many related sequences are being isolated on the basis of the conserved motifs e.g. zinc fingers and tyrosine kinases.

In the absence of obvious similarities with known genes inter-species comparison can yield important information. In the case of the tumour suppressor gene p53, the cloning of a range of orthologs allowed comparison with the human and mouse genes revealing a distinctive pattern of conserved and non-conserved regions (Soussi et al., 1990). Subsequent analysis of mutant p53 in tumours has shown clustering of nucleotide mutations resulting in amino acid changes in the evolutionarily conserved regions (Levine 1992).

With the isolation of more orthologs and paralogs for comparative analysis the study of chromosome evolution is becoming more feasible. The chromosomes of vertebrates are surprisingly variable, considering how well the actual genes themselves have been conserved. The way in which some of these changes have occurred can be recognised by defining homologous syntenic regions conserved between species. In vertebrate evolution several tetraploidisations have been known to occur and these have been followed by reorganisation to regain diploidy. The way in which this occurred may tell us about chromosome structure and behaviour (Lundin 1993).
1.14.2 Functional analysis

The genes present in organisms can to some extent be grouped by their function into classes such as metabolic, cell cycle control, and developmental specification. The conservation of a gene and its maintenance through evolution will depend on its function. In the case of the metabolic enzymes which fulfil very basic and necessary functions, many are conserved in both prokaryotes and eukaryotes. Glutamate synthetase is one such gene which duplicated 3,500 Myra (Kumada et al., 1993) and GSI or GSII is found in every kingdom. The regulation of the eukaryotic cell cycle is another highly conserved function, the human homolog of yeast *Schizosaccharomyces pombe cdc2* has been found to be capable of complementing *Schizosaccharomyces pombe* lacking functional *cdc2* (Lee and Nurse 1987). Within the vertebrates cell regulatory and tumorigenesis mechanisms can be conserved, as was recently demonstrated in that zebrafish *c-myc* was capable of co-operating with mammalian *H-ras* to transform mammalian cells (Schreiber-Agus et al., 1993).

The regulation of development has evolved with the advent of Metazoa. In recent years much has been learnt about the genetic control of development of the fruit fly *Drosophila*. One of the major findings is that many of the *Drosophila* developmental genes have homologs in the vertebrates which also have roles in development. This is the case for *Notch* (Bierkamp and Campos-Ortega 1993), the *Wnt* family (Sidow 1992), the *hedgehog* family (Ingham 1994), and the *HOM-C* and *Hox* gene clusters (Holland 1990). Most of the genes seem to be involved in position specification in the embryo as opposed to specifying specific structures. The resemblances between phyla in *Hox* gene expression has lead to the proposal of the 'zootype' stage where the animal body plan is being set up and can be recognised by characteristic gene expression patterns (Slack et al., 1993). In many cases this coincides with the phylotypic stage which was based on morphological similarities.

*WT1* is a gene that does not seem to be involved in pattern specification but rather in subsequent differentiation of a predetermined region. The evolution of the vertebrate genitourinary system, which is effectively the modification of a basic plan, is quite well understood. The similarities in nephrogenesis and nephroblastoma throughout the vertebrates would imply that the investigation of the anamniote kidney could be applied to the understanding of the development of the amniote kidney and Wilms' tumour.
1.14.3 Animal models

The mouse is currently the favourite animal model for understanding human development. This has become a much more powerful system since the advent of transgenic techniques. For developmental genes, the disadvantage of viviparity, making embryonic stages not particularly accessible has yet to be overcome. In the study of Wilms' tumour, the mouse has the added disadvantage of a very low level of incidence of this tumour. The more user friendly free-living embryos of anamniotes have been exploited. In particular much information has been derived from Xenopus laevis on control of early development, and now new contributions to knowledge are coming from zebrafish (Brachydanio rerio) systems which are genetically more manipulable (reviewed in Strahle and Ingham 1992) especially with the advent of transgenic fish. All of these different systems have advantages and disadvantages, their combined use can allow different questions to be asked and the answers to be compared. The most recent example of this are the studies of chick, mouse and zebrafish hedgehog paralogs (Echelard et al., 1993, Krauss et al., 1993, Riddle et al., 1993).

1.15 Conclusion and aims of the project

Since its isolation in 1990 much has been learnt about WTI but it is far from being fully understood. The specific DNA binding activity and nuclear localisation of WTI is consistent with a role as a transcription factor, but true binding sites for both +KTS and -KTS isoforms remain to be defined. The most often observed effect on transcription by WTI was transrepression but this is dependent on the number and positioning of binding sites and on cell type. The need for multiple binding sites for repression to occur may be due to WTI interacting with itself and it maybe this interaction with which mutant WTI interferes. The cell type specificity and the interaction with p53 indicates that transrepression is not an intrinsic property of WTI but depends on the proteins with which it interacts. Therefore in different cell types in vivo and as cells differentiate the action of WTI may change from activator to repressor of genes or vice versa. Most of the analyses have been performed by transient transfection assays which result in very high non-physiological expression levels and it has been shown that the results obtained for WTI were very sensitive to the systems used so that its in vivo actions may be very different from those observed in vitro.
The histological and mutational observations described above suggest that loss of normal WT1 function would affect normal differentiation, specifically disturbing the mesenchymal-epithelial transition of nephrogenesis. Disturbances of this type are thought to cause the nephrogenic rests, which would then be open to neoplastic transformation. From in vitro studies of WT1 function, loss of regulation of growth factor expression and loss of WT1 autoregulation are candidate mechanisms at the protein level. If the cells cannot pass through the WT1 expressing stage, i.e. the induced blastema and differentiating nephrons, this would explain the persistent expression of the gene as seen in some tumours. In other cases the cells fail to form nephrons and take other developmental pathways where WT1 is not expressed.

It seems that some cases of Wilms' tumour probably do involve the loss of both functional copies of WT1, but it is also clear that there is at least a second predisposing locus at 11p15.5 which can either act alone or in concert with WT1 malfunction leading to tumorigenesis. Also there is some evidence for WT1 acting in a dominant manner; as an oncogene. It is not yet possible to correlate tumour type, WT1 expression and presence and type of WT1 mutations. The only correlation that has been found so far is between BWS (and therefore WT2) and perilobar rest associated tumours, and between WAGR and DDS (and therefore WT1) and intralobar rest associated tumours.

In order to add to the understanding of how WT1 is acting in development and tumorigenesis it was proposed to isolate WT1 from as many relevant developmental systems as possible. First, this would produce a structural comparison which can be a powerful tool in the dissection of protein function. Second, the comparison of WT1 expression patterns would demonstrate if WT1 has a conserved function in genitourinary development and therefore whether the study of WT1 in different classes of vertebrates would be relevant to the study of the function of WT1 in humans.
Chapter 2

Materials and Methods
2.1 Bacterial manipulation

2.1.1 Genotypes

2.1.1.1 Cell strains

**JM101** $\text{supE, thi-1, } \Delta(lac-proAB), [F' \text{ traD36, proAB, lacI}\Delta M15]$ (Yanish-Perron et al., 1985).

**XL1-Blue** $\text{recA1, endA1, gyrA96, } \text{ thi-1, hsdR17, supE44, relA1, lac, } [F' \text{ proAB, lacI}\Delta M15, Tn10(\text{tet}^R)]$ (Stratagene).

**Y1090** $\Delta(lac)U169, \Delta(\text{lon})?, \text{araD139, } \text{strA, supF, mcrA, trpC22::Tn10(}\text{tet}^R), (pMC9; tet'ampR)$ (Young and Davis 1983a).

**SURE** $\text{e14-(mcrA), } \Delta(\text{mcrCB-hsdSMR-mrr})171, \text{ sbcC, recB, recJ, umuC::Tn5 (kan')}, \text{ UVrC, supE44, lac, gyrA96, relA1, thi-1, endA1 } [F' \text{ proAB, lacI}\Delta M15, Tn10(\text{tet}^R)]$ (Stratagene).

2.1.1.2 Bacteriophage

**λZAP** $\lambda\text{shhi}\lambda^O \text{chiA131 (Tamp ColE1 ori lacZ T3 promoter-polycloning site- T}7\text{promoter I) srI}\lambda^3\text{ cIts857 srI}\lambda^4\text{ o nin50 srI}\lambda^5\text{0 Sam100.}$

**λZAPII** as for λZAP but no Sam100 mutation (Short et al., 1988).

**λgt11** $\lambda\text{lac5 }\Delta\text{shndIII}\lambda2\text{-3 srI}\lambda^3\text{ cIts857 srI}\lambda^4\text{ o nin5 srI}\lambda^5\text{0 Sam100}$ (Young and Davis 1983b).

2.1.2 Electro-transformation of *E. coli*

2.1.2.1 Preparation of electro-competent cells

Bacterial cells were streaked out onto an L-agar plate from a -20°C glycerol stock. 10ml of L-broth were inoculated with a single colony and grown overnight in a shaking incubator at 37°C. This was used to inoculate 1 litre of L-broth. Cells were grown until they were in log phase, this was estimated for XL1-Blues or JM101s to be at an OD$_{600}$ of approximately 0.8.

Cells were harvested by chilling the culture on ice as rapidly as possible and leaving on ice for 30 minutes. This was spun at 7,000rpm for 15 minutes and the cell pellet resuspended in 1 litre cold sterile dH$_2$O. This was centrifuged three times more and the cells concentrated down by resuspension in 500ml dH$_2$O, 20ml 10% glycerol

---

1Recipes for media and solutions are at the end of the methods section.
and 2-3ml 10% glycerol respectively to give an approximate final cell density of 3x10^{10} cells/ml. Aliquots were frozen in liquid N$_2$ and stored at -70°C.

### 2.1.2.2 Electro-transformation

Salt was removed from the DNA to be electroporated by dialysis for 15 minutes on a 0.025μm type VM Millipore filter floating on dH$_2$O. Competent cells were thawed, and 40μl added to a cold 0.4 cm electroporation cuvette containing 1-5μl ligation mixture or 1ng plasmid DNA and this was allowed to stand on ice for 1 minute. The cells were pulsed on a BRL Gene Pulser$^TM$ set at 25μF, 2.5kV and 200Ω. 1ml of cold L-broth was added immediately afterwards and the cells were allowed to recover at 37°C for 1 hour with shaking before 100-150μl was plated on agar containing the appropriate selection.

### 2.1.3 Selection of cells harbouring recombinant plasmids

All the vectors used in this study carry ampicillin resistance. Therefore cells harbouring plasmids were selected for after transformation by the presence of 100μg/ml ampicillin in the agar. The pBluescript vectors used contain a partial lacZ gene and the host cells used contain an F' episome harbouring the rest of the gene. LacZ expression can be induced by addition of 40μg/ml IPTG (isopropylthiogalactose) to the medium, the two parts of the gene are then capable of metabolising X-gal (5-bromo-4-chloro-3-indolyl-galactose) which is also added to the medium at 40μg/ml. The product formed is a blue precipitate. If an insert is present in the vector the lacZ gene is disrupted and the colonies are white. White colonies were streaked out onto a fresh agar plate.

### 2.1.4 Bacterial colony lifts

After streaking out, colonies were allowed to grow overnight, the plates were then chilled for at least two hours at 4°C. Gridded nitrocellulose circles (Schleicher & Schuell) were gently placed over the plates and left until the whole circle had absorbed moisture, approximately two minutes. The filters were treated by placing them DNA side up on to Whatmann 3MM paper soaked with denaturing solution for 5 minutes, and then on to paper soaked in neutralising solution for 5 minutes. These were washed briefly (< 1 minute) in 2xSSC. (The cell debris and agar were removed in the 2xSSC wash by rubbing with gloved fingers.) Excess moisture was then
removed with blotting paper. DNA was fixed to the nitrocellulose by baking for 2 hours at 80°C under vacuum. The agar plates were placed at 37°C overnight to allow the colonies to recover.

**2.1.5 Small scale preparation of plasmid DNA**

**2.1.5.1 Method 1**

A bacterial culture was prepared by the inoculation of 1-1.5ml of L-Broth or Terrific Broth containing selection with cells picked from a single colony. This was incubated overnight at 37°C in a shaker. The cells were spun down in a microfuge for 1 minute and as much growth medium as possible removed. The pellet was resuspended in 100μl GTE solution, then 200μl of 1% SDS, 0.2mM NaOH were added and the tube inverted several times to mix the solution. This was incubated on ice for 5 minutes, then 150μl 3M5MKAc solution was added, mixed by inversion, and replaced on ice for 5 minutes. The suspension was spun for 5 minutes and the supernatant retained. To this 1 volume of EtOH was added, vortexed and the DNA precipitate spun down for 5 minutes in a microfuge. The pellet was washed in 1ml 70% EtOH and freeze dried. This was resuspended in 20μl TE buffer or sterile dH2O. This method produced high quality DNA which was suitable for DNA sequencing (Jones and Schofield 1990).

**2.1.5.2 Method 2**

The above procedure was used for most of the experiments but in the final year the Magic™ Miniprep purification system was produced by Promega. The main advantage was the increased yield, purity, and speed of the system.

Cells were harvested from 1-3ml of overnight culture as above. The cell pellet was resuspended in 200μl of cell resuspension solution and inverted until the solution cleared. To this was added 200μl cell lysis solution, the solutions were mixed by inverting the tube and then 200μl 2.55M KAc was added and mixed as before. This was spun in a microfuge for 5 minutes and the supernatant retained. 1ml of warmed (25°C) resin (containing 7M guanidine HCl) was added to the supernatant. This was pipetted into a 2ml syringe barrel attached to a Magic™ Column. The solution was pushed though the column and the column washed with 2ml column wash solution. The column was dried by spinning for 20s and leaving at 50-60°C for 5 minutes. 50μl of dH2O at 65-70°C was applied to the column and the DNA was eluted into an
2.1.6 Large scale preparation of plasmid DNA

500ml of Terrific broth containing the appropriate selection were inoculated with bacteria from a single colony. This was incubated overnight at 37°C with shaking. The cells were harvested by centrifugation at 6,000rpm for 5 minutes and the cell pellets drained on ice to remove as much of the growth medium as possible. The cells were resuspended in 18ml of GTE and then 2ml of freshly prepared lysozyme solution (10mg/ml in 10mM Tris-HCl pH8.0) was added. To lyse the cells 40ml 1% SDS, 0.2M NaOH was added and mixed by inversion of the centrifuge bottle. After 10 minutes on ice 30ml 3M KOAc was added, mixed thoroughly, and replaced on ice for 10 minutes. This was centrifuged at 10,000rpm for 30 minutes and the supernatant filtered through muslin. To this 0.6 volume isopropanol was added and incubated at RT for 10 minutes. The DNA precipitate was then spun down at 8,000rpm for 10 minutes at RT. The pellet was drained, washed with 70% EtOH and freeze dried. The pellet was resuspended in 6ml TE buffer, care being taken to avoid shearing of the high molecular weight chromosomal DNA. The volume of the DNA solution was measured and 1g/ml CsCl was added and the volume re measured (the solution was warmed to 30°C to facilitate dissolution of the salt). To this 0.08 volume 10mg/ml ethidium bromide solution was added. This solution was put into Beckman Quick-Seal® Bell Top centrifuge tubes and sealed. Balanced tubes were spun at 100,000rpm in a Beckman TL-100 ultracentrifuge for at least 4 hours. Of the two bands visible in the CsCl gradient the lower plasmid band was removed by piercing the tube with a needle and drawing out the band into a syringe while avoiding the upper chromosomal band.

The ethidium bromide was removed by butan-1-ol extraction until the solution was colourless. The volume of DNA solution was kept constant by adding H2O. To the plasmid DNA, 3 volumes dH2O and 2 volumes EtOH were added and precipitation of the DNA allowed to occur over 30 minutes at -70°C. The DNA precipitate was spun down at 12,000rpm for 10 minutes, washed with 70% EtOH and freeze dried. The DNA was resuspended in 1ml TE buffer and the concentration estimated by measuring OD260nm in a spectrophotometer. This assumes that an OD260nm of 1 is equivalent to 50ng/ml DNA. The ratio of the optical densities at 260nm and 280nm was used to assess the purity of the DNA, as at 280nm any protein
contamination is detected. For an acceptably pure DNA solution a ratio of approximately 2 was expected.

2.1.7 Glycerol stocks of bacteria

Stocks of bacterial cells were produced by setting up an overnight culture as for a small scale plasmid preparation. The cells were grown overnight and then chilled and a 1:1 ratio of 50% filter sterilised glycerol to culture added. Stocks were then stored at -70°C. When bacteria were required a sterile inoculation loop was dipped into the briefly defrosted stock and used to streak an agar plate containing selection, if appropriate. The plate was incubated at 37°C overnight and then stored at 4°C.

2.1.8 Isolation of clones from bacteriophage libraries

2.1.8.1 Preparation of host cells for bacteriophage

25 ml of L-Broth containing 0.2% maltose (filter sterilised) and 10mM MgSO₄ was inoculated with a single colony of bacterial host cells from a fresh agar plate. These were grown overnight with shaking at 30°C. The cells were harvested by centrifuging at 4,000rpm for 10 minutes. The cell pellet was resuspended in 10mM MgSO₄ to give an OD₆₀₀nm of 0.5 for screening or amplification, or 1.0 for in vivo excision. Cells were made fresh whenever possible but it was found that infection was not impaired for cells stored at 4°C for up to 2 weeks.

2.1.8.2 Titering of libraries

Serial dilutions of the lambda phage libraries were made in SM buffer. These were added to 200μl of host cells and incubated at 37°C for 15 minutes to allow infection to occur. 3ml of CY top agarose at 43°C was added, mixed quickly, and poured onto prewarmed, predried 9cm plates containing CY select agar. The top agar was allowed to harden for 10 minutes at RT and then incubated at 37°C overnight. The number of plaques produced were then counted and the number of plaque forming units (pfu) calculated for the undiluted library.
2.1.8.3 Screening libraries

24cm x 24cm plates (Nunc) were sterilised by wiping with ethanol and then exposing to ultraviolet light for a minimum of 1 hour. These were poured with 150 ml CY select agar and allowed to dry. 0.8 ml cells were inoculated with an estimated 1x10^5 phage and incubated at 37°C for 15 minutes. This was then plated as above using 30 ml of CY top agarose and incubated overnight. Typically four plates were used so approximately 4x10^5 phage were screened per experiment.

The plates were chilled at 4°C for at least 2 hours before lifts were taken. Lifts were taken by lowering a 20 cm x 20 cm Hybond N+ filter (Amersham) onto each plate and leaving for two minutes. The position of the filter on the plate was marked by pricking with a needle attached to a syringe containing ink through the filter into the agar to make an asymmetric pattern. The filter was removed with care in order not to disturb the top agar. Duplicate lifts were made, the second filter was left on the agar twice as long as for the first lift. The filters were then processed by immersion, phage side up in:

- denaturing solution for 2 minutes
- neutralising solution for 5 minutes
- washing solution for < 1 minute (0.2M Tris-HCl pH 7.5, 2xSSC)

After removal from the last solution excess moisture was removed on blotting paper. Any top agar attached to the filters was removed by rubbing with gloved fingers at the denaturation and washing stages. The DNA was fixed to the filters by laying them DNA side up on blotting paper soaked in 0.4 M NaOH and then washing them in 5xSSC. The filters were then hybridised in PVC sandwich boxes to an appropriate probe. Positive signals were considered to be those that were present on both duplicates. The filters were lined up with the plates using the needle holes and the positions of positives on the plates identified. These were removed by coring the agar with an inverted yellow tip and put into 1 ml SM buffer. These were left overnight at 4°C or at least 2 hours at RT to allow the phage to diffuse into the buffer. The number of pfu was calculated by estimating 1x10^6 phage per plaque. Serial dilutions were prepared to give 450 and 50 plaques per 9 cm plate. These were plated as for titering and screened as before except gridded nitrocellulose circles (Schleicher & Schuell) were used. This meant that excess top agar was removed in the 2xSSC wash and DNA was fixed to the nitrocellulose by baking for 2 hours at 80°C under vacuum. Screening was repeated a third time so that an isolated and therefore pure
plaque could be cored. This plaque was put into 0.5ml SM buffer containing a drop of CHCl3, and the phage allowed to diffuse out as before.

2.1.8.4 Amplification of phage by the plate lysate method

The isolated phage was amplified by infecting 0.3ml of host cells with approximately 50,000 phage and plating in 3ml CY top agarose onto 9cm CY agar plates as above. These were grown for 8 hours so that the plaques were 1-2mm across. The agar was overlaid with 4-5ml SM buffer and the phage allowed to diffuse into the buffer overnight at 4°C. The SM buffer was collected and spun at 8,000rpm for 10 minutes to remove cell debris. To the supernatant 0.3% CHCl3 was added and in addition 7% DMSO was added to 1ml for a -70°C stock.

λZAP and λZAPII were plated out on SURE or XL1-Blue cells as above, and λgt11 on Y1090 cells as above except 50μg/ml ampicillin was included in the medium. The quality of the agar was found to be variable, and L-agar was sometimes substituted for CY(select) agar.

2.1.8.5 In vivo excision of λZAP phagemid

λZAP phage have been designed so that a plasmid (plasmid) can be excised from the phage and recircularised in vivo. This is due to the inclusion at separate positions in the phage of the sites of initiation and termination of the +strand origin of replication of the f1 bacteriophage which are adjacent to each other in the f1 phage. F1 (helper) phage are co-infected with λZAP, at the site of initiation the phage DNA is nicked and the +strand replicated as far as the terminator sequence and then the new strand is circularised. This is then packaged and secreted. When this is infected into new host cells the second strand is synthesised. The plasmid thus obtained is pBluescriptSK-.

1μl amplified phage isolate and 1μl R408 helper phage (>1x10^6 pfu/ml Russel et al., 1986) were added to 200μl host cells and 200μl SM buffer. The mixture was incubated for 15 minutes at 37°C and then 5ml Terrific broth was added and this shaken at 37°C for 3 hours. The cells were killed by incubation at 70°C for 15 minutes, then the cell debris were spun down at 6,000rpm for 5 minutes. 10μl or 10μl of a 100x dilution of the phagemid stock was added to 0.2ml host cells and incubated at 37°C for 15 minutes. 1-100μl was then spread on to L-agar plates containing 100μg/ml ampicillin and grown overnight at 37°C. Plasmids were then analysed.
To avoid contamination with the f1 phage, DNA was prepared by small scale plasmid preparation from colonies and transformed into electro-competent cells.

2.1.8.6 Purification of λgt11 DNA using LambdaSorb\textsuperscript{TM} (Promega)

Phage DNA is purified using antibodies to the phage particles which are added to the crude plate lysate. This initial purification eliminates purification steps needed at a later stage in other protocols.

1/100th volume of phage adsorbent was added to the plate lysate and incubated on ice for 30 minutes and then spun at 7,500rpm for 10 minutes. The pellet was resuspended in SM buffer (1/10th the original lysate volume), spun in a microfuge for 1.5 minutes and resuspended in 10mM Tris-HCl pH 7.8, 10mM EDTA (1/20th original lysate volume). This was heated for 15 minutes at 70°C. The EDTA chelates Mg\textsuperscript{2+} ions which are necessary for the integrity of the phage coat, so the phage DNA is released. The phage solution was then spun for 2 minutes down to remove the absorbent and two phenol/chloroform extractions and one chloroform extraction were performed to remove the phage proteins. The DNA was ethanol precipitated with 0.5 volume 5M NH\textsubscript{4}Ac and two volumes EtOH and incubation at -20°C for 30 minutes. The DNA was washed with 70% EtOH and resuspended in 20-50µl TE buffer.

2.2 DNA/RNA manipulations

2.2.1 DNA preparation from tissue

The tissue (approximately 1cm\textsuperscript{2}) was chopped as finely as possible and homogenised. To this 10ml 10mM Tris-HCl pH7.8, 150mM NaCl, 10mM EDTA and then 50µg RNase was added. This was incubated at 37°C for 30-60 minutes. 10%120% SDS and 5-10mg proteinase K was added and incubated overnight at 37°C. This was purified by a phenol, a phenol:chloroform, and a chloroform extraction (chloroform contained 1 volume octan-2-ol for every 24 volumes chloroform). The DNA was precipitated by adding 0.5 volume 7.5M NH\textsubscript{4}Ac and 2 volumes of EtOH. The DNA was spooled out onto a sealed Pasteur pipette and air dried and resuspended in 8ml TE. The DNA concentration was measured as for plasmid DNA.
2.2.2 RNA preparation from tissue or cell lines

Tissue was dissected as quickly as possible in sterile conditions in PBS (phosphate buffered saline). Cell pellets from cell lines were produced by trypsinisation of the cell cultures (to release attached cells if necessary) and spinning down the resulting cell suspension and then washing the pellet.

The sample was added to 0.6ml 6M Urea, 3M LiCl in an eppendorf tube (approx. 5x10⁶ cells) and homogenised on ice using a syringe and progressively smaller gauges of needle as far as 25. The homogenate was sonicated using an amplitude of 22μ for 3x20s on ice and then incubated overnight at 4°C. The precipitated RNA was spun down for 15 minutes in a microfuge at 4°C. The pellet was resuspended in 0.5ml LiCl/Urea and incubated 2-3hours at 4°C and spun as before. The pellet was resuspended in 0.4ml 10mM Tris-HCl pH8.0, 0.5% SDS, 20mg proteinase K added and incubated at 37°C for 30 minutes. The protein was removed by a phenol, a chloroform (containing 1/25 octan-2-ol) and then two ether extractions. The RNA was precipitated with 2 volumes EtOH and 1/10th volume 3M NaAc pH 5.2. After overnight incubation at -20°C the RNA was spun down and resuspended in DEPC dH₂O. The OD₂₆₀nm was measured and the concentration was then calculated assuming OD₂₆₀nm of 1 is equivalent to an RNA concentration of 40μg/ml. The OD₂₈₀nm was also measured to assess the purity of the sample. A ratio of OD₂₆₀nm/OD₂₈₀nm of approximately 2 is expected for a acceptably pure sample.

All solutions to come into contact with the RNA were either treated by adding 1/1000 volume DEPC (diethylpyrocarbonate), leaving for at least 2 hours and then autoclaving, or were made up using DEPC treated dH₂O (Tris and urea containing solutions).

2.2.3 Agarose gel electrophoresis

2.2.3.1 DNA electrophoresis

DNA fragments were resolved on agarose gels, the agarose concentration being optimised for maximum resolution at the expected sizes of fragments. For concentrations 0.7-1.2g/l agarose was used but for higher concentrations a mixture of agarose and NuSieve GTG low melting point agarose.
For analytical gels the running buffer used was 1xTBE. Routine analysis was performed using a BRL Horizon 58™ minigel system at 100V, or 60V for low melting point agarose.

For preparative gels and genomic gels, the running buffer was 1xTAE. Preparative gels also used the BRL Horizon 58™ minigel system and genomic gels on maxi (in house made) gel tanks. Preparative gels were run at 60V and genomic gels at 35V overnight.

6x glycerol loading buffer was added to samples before loading. DNA standards were loaded either, λDNA/HindIII fragments, ΦX174/HaeIII fragments, or 1kb DNA ladder (Promega). The DNA was visualised by adding one drop of ethidium bromide to agarose solution before the gel was cast. The DNA-ethidium complex then fluoresced under UV light.

2.2.3.2 RNA electrophoresis

Total RNA was resolved by electrophoresis in 1% agarose gels containing 2.2M formaldehyde. Either midi or maxi gel tanks were used. Before loading the RNA was mixed in a 1:1 ratio with formaldehyde sample buffer (1 volume 10xgel running buffer, 2 volumes formamide, 1.2 volumes formaldehyde) and heated to 65°C for 5-10 minutes and the chilled on ice to denature any secondary structure. To this 1 volume of formamide loading dye was added and 1μl 10% ethidium bromide. 0.24-9.5kb RNA markers (Gibco BRL) were treated in the same way and loaded as size standards. Gels were run in 1xgel running buffer initially for 10 minutes at 100V and then overnight at 45V.

2.2.4 Vacuum transfer of DNA and RNA

Vacuum transfer of DNA and RNA on to nylon filters was performed using the LKB vacuum blotting apparatus. The apparatus was set up according to the manufacture's instruction. A plastic mask was placed over the porous support with a window cut out approx. 5mm smaller than the gel. The filter was cut slightly smaller than this window and carefully placed within it. The gel was placed over the filter so that no gaps between the gel and mask are present and no air bubbles were trapped between the filter and gel. The top frame was screwed down and vacuum pump started and the tightness of the gel-mask seal checked.
DNA was depurinated by the addition of 0.25M HCl to just cover the surface of the gel. After 4 minutes this was replaced with denaturing solution. After a further 3 minutes the denaturing solution was removed and replaced with alkaline transfer buffer. RNA was transferred using 10xSSC. Depending on the size of the gel, transfer took from 35-60 minutes. That the transfer was successful was verified by restaining the gel with ethidium bromide and inspecting the filter under UV light. The filters used were Hybond N or N+ (Amersham). In the case of the positively charged nylon the DNA was already fixed to the filter by alkali, otherwise the RNA or DNA was fixed using a Stratalinker (Stratagene) which gives a fixed UV exposure.

2.2.5 Hybridisation

Filters onto which DNA or RNA had been fixed were wet with 2xSSC. Filters, separated by a nylon mesh, were rolled up and uncurled in Hybaid hybridisation bottles containing 2xSSC, leaving as few air bubbles as possible. The DNA was on the side not contacting the glass. 2xSSC was replaced with Church and Gilbert’s buffer, prewarmed to the hybridisation temperature. Bottles were put into a Hybaid rotisserie oven and filters were prehybridised for at least 1 hour. The probe was incubated at 100°C for 10 minutes and immediately added to the bottles. Approximately 5x10⁵ counts/ml hybridisation mix were used when hybridising to cloned DNA or PCR products, and 1x10⁶ counts/ml hybridisation mix for genomic DNA. Library primary screen filters were hybridised in sandwich boxes in Perspex containers. These were treated in the same way as the bottles except that they were incubated in a shaking oven. After overnight hybridisation the solution was removed and the filters washed with the appropriate concentrations of SSC solution containing 0.1%SDS. Following washing, filters were briefly dried on blotting paper, then wrapped in Saran wrap.

Hybridisation conditions were determined by whether high or low stringency was required, depending on the level of identity of the probe to the target DNA. High stringency hybridisation and washes were at 68°C, with 0.5xSSC and 0.1xSSC washes. Reduced stringency was variable, in general hybridisation and washing took place at 58°C with washes of 3xSSC and 1xSSC, unless indicated in the text.

---

²Counts are in dpm (degenerations per minute) and were calculated using the formula:
Counts in dpm = μCi ³²P added to labelling reaction x % incorporation of label x 2.2x10⁴
2.2.6 Purification of DNA

Purification of DNA away from modification enzymes (and unused primers and nucleotides in the case of PCR) was performed either by phenol:chloroform extraction or GENECLEAN® II (Bio101).

2.2.6.1 Phenol:chloroform extraction

The reaction was made up to at least 100μl by addition of dH2O. To this an equal volume of water saturated phenol was added. The phases were mixed thoroughly by vortexing and then separated by spinning in a microfuge for 1 minute. The top phase was then removed, avoiding any protein at the interface, and retained. This was repeated and then again with a chloroform:octan-2-ol (24:1) mixture. The DNA was then ethanol precipitated.

2.2.6.2 GENECLEAN® II (Bio 101)

To the reaction mixture 3 volumes 6M NaI were added and then 5μl glassmilk (suspension of silica matrix in H2O). If more than 5μg of DNA was being purified an extra 1μl glassmilk per 0.5μg extra DNA was added. This was left on ice for 5-10 minutes. The glassmilk was spun down for 5s in a microfuge and the NaI solution removed, this was repeated to remove any traces of NaI. The glassmilk was resuspended in 200μl NEW wash, pelleted by a 10s spin and the supernatant removed. This was repeated twice but for the final wash 400μl NEW wash was used. Any traces of ethanol from the NEW wash were removed by briefly (1 minute) freeze drying. To elute the DNA the pellet was resuspended in 25μl of H2O at 55°C and incubated at 55°C for 5 minutes. The glassmilk was spun down for 1 minute and the supernatant removed to another eppendorf. This was repeated and the two elutes combined. This was spun briefly (30s) before use to make sure that any glassmilk accidentally carried over was pelleted.

2.2.6.3 Purification of a specific DNA fragment

If a specific fragment was desired a preparative agarose gel was run and the desired band excised with a scalpel. The bands were purified away from the agarose by either GENECLEAN® or Spin-X.

2.2.6.3.1. GENECLEAN® II (Bio 101)

The gel slice was melted at 55°C in 3 volumes NaI and purified as above.
2.2.6.3.2. Spin-X

The agarose slice was placed in the insert of a Spin-X tube (Costar). On spinning in a microfuge the DNA moved though the 0.22\textmu m cellulose acetate filter at the bottom of the insert leaving the agarose behind. The tube was spun twice for 15 minutes, the agarose was redispersed and about 50\textmu l TE buffer added between spins. The DNA in the elute was precipitated by adding 1/10th volume 3M NaAc pH5.2 and 2 volumes EtOH. This was incubated at -20^\circ C for at least 30 minutes, the precipitate was spun down, dried, and resuspended in 20\textmu l.

2.2.6.4 "Double GENECLEAN®" protocol for cloning PCR products

This is a strategy for cloning PCR products which involves firstly GENECLEANing directly or after purification on an agarose gel, then modification of the PCR ends and a second GENECLEAN. The whole PCR reaction was GENECLEANed using 25\textmu l glassmilk and eluting in 20\textmu l dH\textsubscript{2}O. The second GENECLEAN used 10\textmu l glassmilk and eluted in 25\textmu l dH\textsubscript{2}O.

2.2.7 Quantification of DNA fragments

The amount of a DNA fragment isolated was quantified by running an analytical agarose gel with 1-2\textmu l of the DNA solution. In parallel, a range of known amounts of size markers were run at approximately the same concentration as the expected DNA concentration. The amount of DNA in the sample was estimated by comparison with the brightness of a size marker band of approximately the same size.

2.2.8 Enzyme modification of DNA and RNA

2.2.8.1 Restriction digest of plasmid DNA

Typically 1\mu g of DNA was digested per 10\mu l reaction volume with 1 unit of restriction enzyme (Boehringer Mannheim), in the restriction buffer recommended by the manufacturer and at the appropriate temperature. For most enzymes the appropriate temperature was 37^\circ C. If the plasmid DNA had been prepared by the small scale plasmid preparation method 4mM spermidine was included in the reaction and 3\mu g RNase A (DNase free). The cutting of DNA was analysed using agarose gel electrophoresis.
For analytical digestion approximately 0.5-1μg DNA was digested for about 1 hour.

For preparative digestion approximately 10-20μg DNA was digested overnight. After digestion the enzyme was heat inactivated if appropriate.

2.2.8.2 Restriction digests of genomic DNA

Digests were set up containing 10μg mammalian DNA or the genomic equivalent as the size of the genome varied for the classes used. 

10μg mammalian DNA or the genomic equivalent,

5mM spermidine,

1x restriction enzyme buffer,

3U restriction enzyme per 1μg DNA,

dH₂O to 50μl.

To ensure even digestion of the DNA, the reactions were set up minus the enzyme and left at 4°C for 1-2 hours occasionally stirring with a yellow pipette tip. The enzyme was then added, mixed thoroughly and the reactions left for another hour at 4°C and then at 37°C overnight. To check that the digestion was complete 0.5μg was run on a minigel before loading the rest of the sample on a maxi gel.

2.2.8.3 Production of radiolabelled DNA probes

2.2.8.3.1 Random primed labelling

The Boehringer Mannheim Random primed DNA labelling kit was used, utilising a technique developed by Feinberg and Vogelstein (1983). DNA fragments to be labelled were produced by restriction digest and purified using agarose gel electrophoresis and Spin-X.

25-30ng DNA in 9μl dH₂O was boiled for 5 minutes and immediately chilled on ice. To this 1μl 0.5mM dATP, dGTP, dTTP, 3μl 10xreaction buffer (containing 100μl hexanucleotide mixture), 5μl [α-³²P]dCTP (3000Ci/mmol, Amersham) and 2U Klenow enzyme was added. This was incubated overnight at RT or for 1 hour at 37°C. Incorporation of the label was assessed by precipitation of 0.5μl reaction mix on a GF/B Whatmann filter with approximately 3x15ml 5% TCA, where incorporation of 30% or more had occurred the labelled DNA was used as a probe.
Unincorporated nucleotides were removed by purification on a nick translation column (Pharmacia) containing Sephadex® G-50. The DNA eluted in 400μl TE buffer and the unincorporated nucleotides were retained on the column.

2.2.8.3.2 Specific primed labelling

This was used for labelling of the genomic zebrafish probe used for screening cDNA libraries. Random priming was not used due to the small size of the probe, instead the PCR primer C229 which was at the 5' end of the probe was used to prime the extension using Taq polymerase.

25ng DNA was denatured by boiling for 5 minutes in 6.8μl dH₂O and then chilled on ice. To this 3.2μl 10xTaq polymerase buffer (Promega), 1μl 0.5mM dATP, dGTP, dTTP, 1μl 0.1mg/ml primer, 5μl [α-³²P]dCTP, 1μl Taq polymerase (ABI) was added. This was then subjected to three rounds of; 72°C 25 minutes, 94°C 2 minutes, 55°C 1 minute, and a final extension of 72°C 30 minutes, using a Hybaid Omnigene thermal cycler. The incorporation of radiolabel was checked and the probe purified as for random primed probes.

2.2.8.4 Preparation of riboprobes

This was performed using the Boehringer Mannheim DIG RNA labelling kit, utilising the T7 and T3 promoter sites at either side of the cloning site.

The transcription reaction was set up by adding,

1μg of linearised DNA purified by phenol:chloroform extraction,
4μl 10x transcription buffer,
4μl 10mM dATP, dCTP, dGTP, 6.5mM dUTP, 3.4mM DIG-UTP,
1μl 1M DTT,
2μl RNase inhibitor,
2μl T7 or T3 RNA polymerase,
redistilled dH₂O to give total volume of 40μl.

This was incubated for 1 hour at 37°C whereupon a second dose of enzyme was added and the incubation repeated. To remove the DNA template, 2μl (20U) DNaseI and 1μl 10μg/μl tRNA was added and incubated at 37°C for 15 minutes. The reaction was stopped with 4μl 0.2mM EDTA pH8.0 and the RNA precipitated by adding 5μl 4M LiCl and 150μl EtOH. This was left overnight at -20°C, then spun at
4°C for 30 minutes. The RNA pellet was washed three times with 70% EtOH at -20°C and freeze dried. The RNA was resuspended in 20μl DEPC dH₂O and 1μl RNase inhibitor added.

The amount of riboprobe produced was assessed by comparison with the labelled RNA supplied with the kit, using the Boehringer Mannheim DIG Nucleic Acid Detection kit. Serial dilutions of the probe and standard were dot blotted onto Hybond N and fixed by UV exposure. The filters were then washed in buffer 1 for 1 minute at RT on a shaker, then in buffer 2 for 30 minutes. This was followed by another brief wash in buffer 1 and then filters were soaked in buffer 1 containing 1/5000 dilution of the anti-DIG-alkaline phosphatase conjugate for 30 minutes. The unbound antibody was removed by washing twice for 15 minutes in buffer 1, then equilibrated in buffer 3 for 2 minutes. This was replaced with buffer 3 containing 4.5μl NBT solution and 3.5μl BCIP (X-phosphate) and incubated in the dark until the dots became visible (about 20 minutes). The reaction was stopped by washing in buffer 4. In general a yield between 0.1-1μl/μg of RNA was obtained The probe was denatured prior to use in 20μl DEPC dH₂O at 80°C for 5 minutes and cooled on ice.

### 2.2.8.5 cDNA synthesis

This was performed using the Boehringer Mannheim cDNA synthesis kit (Gubler and Hoffman, 1983). The first DNA strand was synthesised by adding,

- 2μl buffer I,
- 0.5μl (12.5U) RNase inhibitor,
- 1μl 40mM dNTP's,
- 1μl 10ng/ml random hexamers,
- 1μl (20U) AMV-Reverse Transcriptase,
- 2μg of total RNA,
- redistilled dH₂O to a total volume 10μl.

Incubation was at 42°C for 1 hour.

The second strand was synthesised by adding,

- 20μl buffer II,
- 0.5μl RNaseH,
2.5µl (12.5U) *E. coli* DNA polymerase I,  
7µl redistilled dH₂O,  
and incubating at 16°C for 1 hour, RT for 1 hour and 65°C for 10 minutes.

### 2.2.8.6 Modification of ragged ends of PCR products

For blunt ended cloning the non template directed extra base added by the Taq polymerase to the 3’ end of the PCR product was chewed back. The reaction was set up as follows:

- One half of a GENE CLEANed PCR reaction,  
- 10µl 10x *E. coli* DNA polymerase I buffer,  
- 1µl 100mM rATP,  
- 10U T4 polynucleotide kinase,  
- 10U *E. coli* DNA polymerase I (BRL),  
- total volume of 100µl.

This was incubated for 1 hour at 37°C and the reaction stopped by the addition of 1µl 0.5M EDTA pH8.0.

### 2.2.8.7 Ligation of DNA fragments into vectors

Initially the pBluescript vector, pSKII+ (Stratagene) was used. This is a derivative of pUC19. It has a high copy number in *E. coli* and contains a multiple cloning site (MCS) with unique restriction sites and the T7 and T3 promoters. It also has a partial *lacZ* gene and so blue/white selection to detect the presence of an insert can be used.

#### 2.2.8.7.1 Cohesive end ligation

The DNA fragment and the vector were digested with the appropriate enzymes and gel purified. The vector and the insert were mixed at a molar ratio of about 3:1 (about 100ng DNA) in a volume of 7µl and warmed for 5 minutes at 45°C and then chilled on ice. 1µl of 10x ligation buffer, 1µl rATP, 1µl (1U) T4 ligase were added and incubated overnight at 16°C.
2.2.8.7.2 Blunt ended ligation of PCR products

pSKII+ was cut with EcoRV in the MCS to produce blunt ends. The PCR product was purified using the "Double GENE CLEAN®" protocol, ragged ends were chewed back to produce blunt ends. 40ng of vector and insert were added to 1μl of 10x ligation buffer, 1μl rATP, 1μl (1U) T4 ligase and incubated at least overnight at 16°C.

2.2.8.7.3 Turbo cloning of PCR products

This is a very fast and non size selective method for cloning (Boyd 1993), and is explained in Chapter 4. pBS:lox is a manipulated version of pBluescribe (pBS) (Stratagene). A recognition site (lox site) for the Cre enzyme has been introduced at the PvuII site 5' to the MCS.

The ends of the GENE CLEANed PCR products were made flush (section 2.2.8.6) and the modifying enzymes were inactivated by treatment at 75°C for 5 minutes, and the DNA was desalted by dialysis on a Millipore filter (section 2.1.2.2). The ligation mixture was set up;

- 7μl (10ng) PCR product,
- 1μl (100ng) Smal cut and dephosphorylated pBS:lox (gift A.C. Boyd),
- 1.6μl 10x buffer,
- 6μl 40%(w/v) PEG 6000,
- 0.4μl (0.4U) T4 ligase.

This was incubated at 20°C for 15 minutes, the ligase was then inactivated by 10 minutes at 65°C. 64μl M restriction enzyme buffer (Boehringer Mannheim) and 0.5μl Cre enzyme (NEN) was added. This was incubated at 30°C for 30 minutes. The enzyme was heat inactivated as above.

2.3 Sequencing DNA

The method used is based on the original method of Sanger et al., (1977) which involved extension of a primer with a polymerase. The DNA produced is labelled by incorporation of [α-35S]dATP and reactions terminated by the incorporation of dideoxynucleotides. The Sequenase® II kit produced by USB was used which includes an engineered version of the T7 polymerase, Sequenase® enzyme. This increases the processivity of the enzyme and therefore increases the readability of
sequence obtained. The primers used were pBluescript primers or those originally designed for PCR.

2.3.1 Double stranded sequencing of plasmid DNA

1-5μg of plasmid DNA in 16μl dH2O was denatured by adding 4μl 1M NaOH, 1mM EDTA and incubating at 65°C for 5 minutes before chilling on ice. The DNA was precipitated by adding 2μl 2M NH4Ac pH4.5, 60μl cold EtOH and leaving on dry ice for 10 minutes. The DNA was pelleted at 4°C for 10 minutes and washed in 1 volume 80% EtOH, then dried.

The annealing mix was made up;

2μl 5xSequenase® buffer,
1μl (100μg/ml) sequencing primer,
7μl DNA pellet resuspended in dH2O

This was heated to 65°C for 2 minutes and then at 37°C for 15 minutes.

The template mix was assembled;

1μl 0.1mM DTT,
2μl 1/5 dilution dGTP labelling mix,
0.5μl [α-35S]dATP (<1000Ci/mmol, NEN),
1μl 1/8th dilution of Sequenase® in enzyme dilution buffer.

The template mix was added to the annealing mix, incubated at RT for 4 minutes and 3.4μl was added to 2.5μl of each termination mix (prewarmed at 37°C). This was incubated for 4 minutes at 37°C, the extension was stopped with 4μl formamide loading buffer. About 3μl of each termination reaction was loaded per lane of a sequencing gel.

2.3.2 Direct sequencing of PCR products

This is an adaption of the Sequenase® method for plasmid DNA sequencing, and involves the use of DMSO to retain the PCR template in a denatured state (Winship 1989).

The annealing mix was made up;
approximately 300ng of gel purified PCR fragment in 6μl dH2O,
1μl 0.5μM primer,
2μl 5xSequenase® buffer,
1μl DMSO.
This was denatured at 100°C for 3 minutes and snap frozen on dry ice/EtOH.
The labelling mix was assembled;
1μl 0.1M DTT,
2μl 1/15th dilution dGTP labelling mix,
0.5μl [α-35S]dATP,
1.5μl Sequenase®.
This was added to the annealing mix as it was thawing. This reaction mix was incubated at RT for 45s and then 3μl added to 2μl of each termination mix (containing 10% DMSO and prewarmed to 37°C). These were incubated at 37°C for 5 minutes and the reactions stopped by adding 4μl formamide dye. 4μl of each termination reaction was loaded per lane of a sequencing gel.

2.3.3 Polyacrylamide gel electrophoresis (PAGE) of sequencing reactions

Prior to using, the glass sequencing plates were washed with soap, dried and one plate coated in dichlorodimethylsilane, to prevent the gel adhering to this plate. The plates were taped together along the sides separated by 0.4cm spacers and gripped together by bulldog clips. The Sequagel system (National Diagnostics) was used for making up acrylamide. Just prior to use the appropriate amounts of concentrate, dilution buffer and running buffer were mixed to give the desired acrylamide concentration. To this was added 1/10 volume 10% ammonium persulphate and 50μl TEMED (N,N,N',N'-tetramethylethylenediamine), this was quickly poured between the plates while the mixture remained liquid. 0.5-0.8% acrylamide was used depending on the resolution required (Sambrock et al., 1989).

Gels were run in 1xTBE using a limiting factor of 30W, the current and voltage being approximately 24mA and 1250V. Gels were typically run between 1.5-5 hours, using bromophenol blue and xylene cyanol dye fronts to estimate the positions of the fragments. A metal plate, clipped to the front of the plates, acted as a heat distributor to avoid 'smiling' of the gel. After the run was completed the gel plates were
separated and the gel fixed in 10% acetic acid, 10% methanol for 15-30 minutes. The gel was then transferred onto 3MM Whatman paper, covered with Saran wrap and dried under vacuum for 1 hour.

2.4 Detection of radioactive signal

2.4.1 Autoradiography

This was the most frequently used method for detecting radioactive signal. Filters hybridised to \( \alpha^{-32}\text{P} \) labelled probes were exposed to Kodak XAR-5 film at \(-70^\circ\text{C}\) with an intensifier screen. In the first instance exposure was overnight except for PCR fragments which were exposed for 1-2 hours. Spots of Glow Juice (IBI) at the edge of the filters were used to line up the filters with the film.

The DNA sequence 'ladders' were visualised by exposure to Kodak XAR-5 film at RT, in the first instance overnight or longer if the signal was faint.

2.4.2 PhosphorImager analysis

This was used towards the end of the project, particularly for genomic DNA and Northern blots. The filter or sequencing gel was placed in a PhosphorImager cassette, and depending on the signal, left for an appropriate amount of time at RT. (1-2 hours for sequencing gels or overnight for blots). The PhosphorImage screen was then analysed by the PhosphorImager using a laser and an image produced and analysed using the ImageQuantTM software (Molecular Dynamics).

2.5 Incubation and analysis of embryos

2.5.1 Rearing of Xenopus embryos

Xenopus eggs were obtained from Professor P.M. Gaze, Department of Zoology, University of Edinburgh. The embryos were kept in clean Xenopus rearing solution and fed on Nettle Powder (Philip Harris Scientific) and Complan (Boots). In order to stop the embryos swimming so that they could be staged (Niewkoop and Faber 1967), the tadpoles were anaesthetised by placing them in a very dilute solution of MS222. Prior to dissection the tadpoles were placed in a more concentrated solution.
2.5.2 Whole mount in situ hybridisation

This protocol is an adaptation by Lesley McInnis of that described by Wilkinson (1992).

2.5.2.1 Incubation and dissection of embryos

Hen eggs were supplied by the AFRC unit at Roslin, East Lothian. These were incubated at 37°C and turned every 24 hours for the first 3 days. Embryos were dissected at RT in sterile PBS. As many extra-embryonic membranes as possible were removed to avoid trapping of the probe/antibody. They were then placed on ice in PBS pH7.2 in Sterilin conical tubes. Embryos were staged according to the tables of normal development (Hamburger and Hamilton 1951).

2.5.2.2 Preparation of embryos

The embryos were treated as follows (all washes were for 10 minutes in 5-10ml):

1) Washed embryos twice in ice cold PBS.

2) Fixed embryos for at least 2 hours in freshly made 4% paraformaldehyde (PFA) in PBS at 4°C.

3) Washed, twice in PBT (PBS+0.1%Tween 20).

4) Dehydrated on ice through 25%, 50%, 75% and 100% (twice) methanol in PBS.

5) The embryos were stored at -20°C until required.

6) Rehydration on ice through the same percentages as above.

7) Washed three times in PBT.

8) Bleached with 6%H2O2 in PBT for 1 hour.

9) Washed three times in PBT.

10) Digested with 10μg/ml proteinase K in PBT for 15 minutes.

11) Washed in 2mg/ml glycine in PBT.

12) Washed twice in PBT.

13) Refixed in fresh 0.2% gluteraldehyde/4%PFA in PBT for 20 minutes.

14) Washed twice in PBT.
2.5.2.3 Hybridisation

Hybridisation of the riboprobe was performed as follows (in a volume of 1ml):

1) Washed in 50% hybridisation buffer: 50% PBT at RT.
2) Washed in hybridisation buffer at RT.
3) Prehybridisation in hybridisation buffer containing 100μg/ml tRNA and 100μg/ml denatured salmon sperm DNA at 70°C for 1 hour.
4) Replaced solution with the same containing 500ng denatured riboprobe.
5) Tubes humidified by placing a 0.5ml eppendorf tube containing hybridisation buffer into the neck of the tube.
6) Hybridisation overnight at 70°C.
7) Washed twice in hybridisation buffer at 70°C for 10 minutes.
8) Washed twice for 5 minutes, and three times for 30 minutes at 65°C in post hybridisation wash.
9) Allowed embryos to cool to RT.
10) Washed in 1xTBST three times.

2.5.2.4 Antibody conjugate binding and staining

1) Diluted anti-DIG alkaline phosphatase conjugate 1 in 500 in 1xTBST, 1% heat inactivated sheep serum (Scottish Laboratories Products) and added 6mg/ml heat inactivated embryo powder from the appropriate species.
2) Preadsorbed antibody 1 hour at 4°C.
3) Blocked embryos with 1xTBST, 10% heat inactivated sheep serum for 1 hour at RT.
4) Replaced blocking solution the same solution containing 1 in 2000 dilution of preadsorbed antibody. Allowed antibody to bind overnight at 4°C.
5) Washed 3 times for 5 minutes and 3-5 times 30-60 minutes in 1xTBST.
6) Equilibrated embryos in fresh alkaline phosphatase buffer by washing 3 times for 10 minutes
7) Replaced solution with alkaline phosphatase buffer containing 4.5μl NBT and 3.5μl BCIP in glass vials.
8) Allowed stain to develop in the dark and examined at regular intervals until a stain appeared.

9) Stopped reaction by rinsing in PBT containing 0.2M EDTA

10) Refixed the embryos in 4%PFA.

2.5.2.5 Photography of embryos

Stained embryos were photographed using a Wild Heerbrugg Photomakroskop M400 and a Wild Leitz WPS Photoautomat using 64 ASA colour film (Fuji).

2.6 Polymerase chain reaction (PCR)

2.6.1 Synthesis and purification of oligonucleotides

PCR primers were synthesised by either D. Chambers or A. Gallacher on an Applied Biosystems 381 A synthesiser. 350µl ammonium hydroxide stock of oligonucleotides was precipitated with 35µl 3M NaAc pH5.5 and 770µl EtOH. This was incubated at -20°C for 30 minutes and the precipitate spun down in a microfuge at 4°C for 15 minutes. The pellet was washed twice in 80% EtOH and then freeze dried. The pellet was resuspended in 200µl TE and the concentration estimated from the OD260nm. An OD260nm of 1 was estimated to be equivalent to 25µg/ml single stranded oligonucleotide.

2.6.2 Reaction conditions

The standard PCR mix contained;

1x Taq polymerase reaction buffer (Promega or nbl (Northumbria biolabs)),
200µM each dNTP,
5µg/ml each PCR primer,
2.5U Taq polymerase (Promega or nbl),
sterile dH2O.

The reactions were set up in 0.5ml eppendorf tubes generally using either 25 or 50 µl reaction mix, and covered with 3-4 drops mineral oil. Amplification was performed by an initial 3 minute denaturation (94°C), and then 30 cycles; 50°C 2 minutes 72°C
1 minute 94°C 1 minute, followed by a final extension of 50°C 1 minute and 72°C 10 minutes.

If hot start PCR was used the reaction mix, minus the enzyme, was heated to 94°C for 3 minutes and the temperature was then held at 72°C while the Taq polymerase was added. The annealing temperature was subject to variation depending on the primers involved and whether they were 100% identical to the template or not. The extension time was increased to 2 minutes or more when long products (>1kb) were expected. For exon 1 PCR 10% glycerol was included to minimise amplification of spurious bands, these probably occurred due to the high GC content of the region and of the primers. Initially reactions were performed using Perkin Elmer Cetus DNA thermal cyclers and later Hybaid Omnigene thermal cyclers.

2.6.3 PCR template

Several different templates were used in the course of this project.

PCR analysis of bacterial colonies was performed by picking the colony and vortexing in 1ml dH2O. The cell suspension was boiled for 1 minute and chilled on ice. 1μl was then used as a template.

For PCR directly from bacteriophage, 5μl phage suspension in SM buffer were placed in the balance dH2O for the PCR reaction and boiled for 5 minutes, the rest of the PCR mix was then added.

If cDNA was used 1/10th-1/5 of the PCR reaction volume was added, equivalent to 25-50ng RNA per 50μl PCR reaction.

For PCR from genomic DNA, 0.3μg genomic DNA was added per 100μl reaction.

2.7 Solutions

Bacteria and bacteriophage media

L-Broth per litre, pH7.2
10g tryptone (Difco),
5g yeast extract (Difco),
10g NaCl,
2.46g MgSO4.

L-agar contains L-both plus 15g agar/litre.
| CY (select) agar per litre, pH7.2 | 10g casamino acids (Difco), 5g yeast extract, 3g NaCl, 10g agar. |
| CY top agarose | contains 6.5g/l agarose in place of agar. |
| Terrific Broth per 800ml | 12g tryptone, 24g yeast extract, 4ml glycerol. |
| KPO₄ buffer | 170mM KH₂PO₄, 720mM K₂HPO₄, 1/10 volume added to Terrific broth. |
| SM buffer | 100mM NaCl, 10mM MgSO₄, 50mM Tris-HCl pH7.5, 0.01% Gelatin. All reagents were sterilised by autoclaving before use. |
| Ampicillin stock solution | 100mg/ml in dH₂O (filter sterilised). |
| IPTG stock solution | 20mg/ml in dH₂O (filter sterilised). |
| X-Galactose | 20mg/ml in dimethylformamide. |
| **DNA manipulation and purification** |
| Denaturing solution | 1.5M NaCl, 0.5M NaOH. |
| Neutralising solution | 1.5m NaCl, 0.5M Tris-HCl pH8. |
| 20xSSC | 3M NaCl, 0.3M Na₃citrate. |
| GTE | 50mM Glucose, 25mM Tris-HCl pH8.0, 10mM EDTA. |
| TE | 10mM Tris-HCl pH7.6, 1mM EDTA pH8.0. |
Magic Minipreps™

Cell Resuspension Solution
50mM Tris-HCl pH7.5,
10mM EDTA,
100mg/ml RNase A.

Cell Lysis Solution
1% SDS,
0.2M NaOH.

Column Wash Solution
100mM NaCl,
10mM Tris-HCl pH7.5,
2.5mM EDTA,
50% EtOH.

Electrophoresis and hybridisation

1xTAE
40mM Tris-Acetate,
1mM EDTA pH8.0.

1xTBE
90mM Tris-Borate,
2mM EDTA pH8.0.

10x RNA gel running buffer
0.4M MOPS (morpholinopropansulphonic acid),
100mM NaAc,
10mM EDTA pH8.0.

6x glycerol loading buffer
0.25% bromophenol blue,
0.25% xylene cyanol FF,
30% glycerol in dH2O.

Formamide loading buffer
95% formamide,
20mM EDTA,
0.05% bromophenol blue,
0.05% xylene cyanol.

Alkali transfer buffer
0.25 NaOH,
1.5M NaCl.

Church and Gilbert's solution
0.5M Na Phosphate pH7.2,
7% SDS.

Enzyme buffers

10x E.coli DNA polymerase I buffer
0.5M Tris-HCl pH7.5,
0.1M MgCl₂,
10mM DTT,
500µg/ml BSA (Fraction V),
200µM dNTPs.
10x T4 Ligation buffer
200mM Tris-HCl pH7.6,
50mM MgCl₂,
50mM DTT,
500μg/ml BSA (Fraction V).

10x Turbo cloning buffer
500mM Tris-HCl pH8,
5mM rATP,
5mM DTE,
50mM MgCl₂.

10x Taq polymerase
buffer (Promega)
1.5mM MgCl₂,
50mM KCl,
10mM Tris-HCl pH 9.0,
1% Triton X-100.

Alkaline phosphatase colour reaction
Buffer 1
0.1M Tris-HCl pH 7.5,
0.15 M NaCl.

Buffer 2
buffer 1 containing 1% blocking reagent.

Buffer 3
0.1M Tris-HCl pH9.5,
0.1M NaCl,
50mM MgCl₂.

Buffer 4
10mM Tris-HCl pH8.0,
1mM EDTA.

NBT solution
75mg/ml nitroblue tetrazolium in dimethylformamide.

BCIP or X-phosphate
solution
50mg/ml toluidinium 5-bromo-4-chloro-3-indolyl-phosphate
in dimethylformamide.

Whole mount in situ hybridisation
Hybridisation buffer
50% deionised formamide,
5xSSC,
50μg/ml Heparin,
0.1% Tween-20.

pH to 5 with 1M citric
acid.

Post hybridisation Wash
2xSSC,
50% formamide,
0.1% Tween-20.
1xTBST  140mM NaCl,  
    3mM KCl,  
    25mM Tris pH7.5,  
    0.1% Tween-20.

Alkaline phosphatase  100mM Tris-HCl pH9.5,  
buffer  100mM NaCl,  
    50mM MgCl₂,  
    0.1% Tween-20.

Heat inactivated embryo  ground, acetone dehydrated, homogenised embryonic tissue,  
powder  heated at 70°C for 30 minutes in 1xTBST

Xenopus rearing  60mM NaCl,  
solution  0.6mM KCl,  
    0.4mM MgSO₄,  
    0.3mM Ca(NO₃)₂,  
    0.8mM Na₂HPO₄,  
    0.15mM KH₂PO₄,  
    2.4mM NaHCO₃.

2.8 Computer analysis of DNA sequences

The University of Wisconsin's Genetics Computer Group have produced an extensive and expanding set of programs for DNA sequence analysis and manipulation (GCG 1991). Routine functions were performed using these programs. In addition sequence comparison of the WTI orthologs was performed using the gap program. The pepplot program was used to draw the protein structure prediction plots, motifs and helicalwheel programs were used to investigate protein structure motifs. The Blast program was used to search the EMBL, GenBank and SwissProt sequence databases using sequences isolated in this study (Altschul et al., 1990, Karlin and Altshul 1990). Clustal V was used for phylogenetic analysis and multiple alignment of the WTI sequences (Higgins and Sharpe 1988, Higgins et al., 1992), and is described in Chapter 5. Signal Scan was used to locate the presence of any possible transcription binding sites (Prestridge 1991). All the above programs were used via the Human Genome Mapping computing facility.

The Coils program (Lupas et al., 1991) and three updates were used to analyse the potential of WTI to form coiled coils. A UNIX version (Yauso Ina 1993) of Li's 1993 program, lwl15, (Li et al., 1985, Li 1993) was used for the calculation of KS and KA to investigate the rate of evolution of WTI.
Chapter 3

Identification of sequences cross hybridising to WTI in the vertebrates
3.1 Introduction

Initially, to determine whether sequences homologous to the \textit{WTI} gene were present in species other than mouse and human, hybridisation of a \textit{WTI} probe to a range of vertebrate DNA's was carried out at low stringency. This technique known as zoo blot hybridisation is often used in gene mapping to determine whether a single copy probe is conserved across species and therefore may correspond to exonic gene sequence (Monaco \textit{et al.}, 1986). It has been used to demonstrate, for known genes, the conservation of homologous sequences in the genomes of different species, in some cases as far as \textit{Drosophila} e.g. \textit{PAX6} (Ton \textit{et al.}, 1991b).

Comparison of nucleotide sequences of mouse and human \textit{WTI} showed 81\% overall identity of the entire cDNA and this increased in the protein coding region to 91\%, and was even higher in the zinc finger region at 95\% (Buckler \textit{et al.}, 1991). This indicated that the \textit{WTI} gene is highly conserved between mouse and human and therefore probably highly conserved within the rest of the mammalian lineage at least.

3.2 Detection of sequences cross hybridising to \textit{WTI} in mammals

In order to determine the presence of \textit{WTI} in other mammalian species apart from human and mouse, genomic DNA was obtained from human (\textit{Homo sapiens}), mouse (\textit{Mus domesticus}), rat (Wistar), guinea pig (\textit{Cavia porcellus}), pig (\textit{Sus scrofa}), sheep (\textit{Ovis aries}), and cow (\textit{Bos taurus}) (gifts R. Hill and A. Archibald). A mouse \textit{Wt1} cDNA clone \textit{pK/S2} (gift Buckler \textit{et al.}) containing all but the first 779bp of exon 1 was used to probe the EcoRI restricted DNA's at reduced stringency. The hybridisation temperature that was used was determined using the formula of 1-1.5°C reduction in melting temperature (Tm) per percentage mismatch of the probe with the target DNA (Sambrook \textit{et al.}, 1989). Washing was carried out in stages and autoradiography performed before the stringency was increased.

Bands were observed in the pig, sheep and cow samples after washing at 58°C in 2xSSC. For human, mouse, rat and guinea pig, bands were seen after washing at 67°C in 0.5xSSC (Figure 3.1).

The strongest band detected in mouse was 8kb and this corresponded with the band seen in C57/B16 mice (Buckler \textit{et al.}, 1991) under high stringency conditions. In human genomic DNA a \textit{WTI} cDNA probe detects 8 EcoRI fragments (13.6, 10.4, 7.2, 6.1, 5.8, 3.7, 3.1, 1.85kb) (Call \textit{et al.}, 1990, Haber \textit{et al.}, 1991). In this
Figure 3.1
Detection of sequences that hybridise to \textit{WTI} in a range of mammalian genomes.

a Ethidium bromide stained agarose gel of EcoRI digested genomic DNA.

b Autoradiographs after hybridisation at reduced stringency (58°C) with mouse \textit{Wtl} cDNA (pK/S2).

Washing conditions at which specific bands were observed are indicated.
experiment the 3.7 and 3.1 kb bands were faint and the 6.1kb band was not detected but all the rest were seen (Figure 3.1b). Therefore it was thought that the conditions used were detecting WTI sequences and so it could be presumed that sequence homologous to WTI were present in rat, guinea pig, pig, sheep and cow genomes (Figure 3.1).

3.3 Detection of sequences cross hybridising to WTI in non mammalian vertebrates

A major problem with low stringency hybridisation with a probe containing a motif is that hybridisation may be occurring to paralogous or related genes also containing this motif, and therefore giving spurious signals. In the case of TFIIIA-like zinc fingers it has been estimated that there are 300-700 finger containing genes in the vertebrate genome (Crossley and Little 1989). In order to avoid cross finger hybridisation a 454bp NcoI restriction digest fragment of WT33 was subcloned (H454). This contained of part of exon 1 and exons 2-4 but not the zinc finger encoding region. WT33 was one of the first human WTI cDNA clones to be isolated, containing all but the 5' 630 bp (Call et al., 1990).

Human, mouse, rat, pig, chick (Gallus domesticus), frog (Xenopus laevis) and trout (Salmo) genomic DNA was digested either with BamHI or EcoRI. The amount of DNA used for the digestion was calculated so that the equivalent number of haploid genomes was used for each species (Table 3). Hybridisation was performed at 55°C with the human probe H454 and washing carried out until bands were observed (2xSSC). These conditions had previously been shown to detect cross species hybridisation for SRY-like sequences in vertebrate genomes (Tiersch et al., 1991). Bands were observed in human, mouse, chick, frog and trout with the BamHI digest (Figure 3.2) and frog with EcoRI digest (data not shown).

3.4 Conclusions

Zoo blotting of vertebrate genomic DNA has demonstrated the presence of sequences capable of cross hybridisation to WTI probes in four of the five vertebrate classes (mammal, bird, amphibian, and fish). These may not be WTI orthologs but closely related sequences, however it was shown that mostly WTI specific bands were being detected in species where the sizes of WTI specific bands were known (Figure 3.1). Secondly the possibility of hybridisation to other zinc finger genes was reduced.
by the use of a probe not containing the zinc finger encoding region. The sequence of
the part of \textit{WTI} sequence used was not similar to any known sequence on searching
of the EMBL and GenBank databases. Therefore these results showed that it was
likely that \textit{WTI}-like genes were present in all vertebrate genomes and gave
encouragement to attempt the cloning of \textit{WTI} from various species.

Table 3 Size of vertebrate genomes.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Nucleotide pairs $\times 10^9$ per haploid genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Homo sapiens}</td>
<td>3.4</td>
</tr>
<tr>
<td>\textit{Mus musculus}</td>
<td>3.36</td>
</tr>
<tr>
<td>\textit{Rattus norvegicus}</td>
<td>2.9</td>
</tr>
<tr>
<td>\textit{Gallus gallus}</td>
<td>1.2</td>
</tr>
<tr>
<td>\textit{Xenopus laevis}</td>
<td>6.2</td>
</tr>
<tr>
<td>Trout</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Figure 3.2
Detection of sequences that hybridise to WTI in a range of vertebrate genomes.

a Ethidium bromide stained gel of BamHI digested genomic DNA

b Autoradiograph after hybridisation at reduced stringency (55°C) with a subclone of human WTI including the region 5' to the zinc fingers (H454).
Chapter 4

Isolation of WTI orthologs
4.1 Isolation by cross species PCR

4.1.1 Introduction

Since it had been shown, by reduced stringency hybridisation, that there were sequences within non mammalian vertebrate genomes which hybridised to a WT1 probe, the cloning of sequences homologous to WT1 within the vertebrates was attempted. The traditional method for the isolation of sequences homologous to a known gene, such as orthologs and paralogs, has been by reducing the stringency of hybridisation when screening libraries. This can be a very time consuming process and often many unrelated sequences are isolated.

In the last few years a major change in methodology has come about with the invention of the Polymerase Chain Reaction (PCR). PCR was developed in 1984 by the Cetus Corporation and is a method of amplification of specific DNA sequences in a cell-free system. PCR abrogates the need for the initial cloning of the DNA prior to amplification (Saiki et al., 1985, Mullis and Faloona 1987). It is based on two oligonucleotide primers flanking the region of interest on opposite strands of the DNA, and cycles of annealing the primers to the template, extension of the primers by a DNA polymerase, and denaturation of the product. The presence of two primers instead of one results in an almost exponential, rather than linear, amplification of the target region (Mullis and Faloona 1987). The speed and sensitivity of PCR has been enhanced by microchip controlled heating blocks, and isolation of a heat stable DNA polymerase (Taq polymerase) from Thermus aquaticus. The simplicity of this technique lends itself to manipulation and therefore a wide range of uses (reviewed by Arnheim and Erlich 1992).

Although PCR was originally developed for situations where the template sequence was known and so complementary primers could be designed (e.g. for the analysis of mutations (Saiki et al., 1985)), it quickly became clear that by lowering the stringency of the reaction a degree of mismatch between the primer and template could be tolerated. This has been exploited to amplify related genes, genes from different species and phyla (Orlandi et al., 1989, Sarker and Sommer 1989, Pang et al., 1990, Sarker et al., 1990, Sommer et al., 1992), or where only the amino acid sequence of the protein is known (Lee et al., 1988). Once part of the target gene has been isolated this can be used as an homologous probe for library screening (Lee et al., 1988).

For the isolation of WT1 orthologs cross species PCR was particularly applicable for several reasons. The speed of the technique makes it the best
Table 4.1 Sources of RNA from which the isolation of *WTI* orthologs was attempted

<table>
<thead>
<tr>
<th>Species</th>
<th>Taxon</th>
<th>Stage/Age</th>
<th>Tissue</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse (BALB/c)</td>
<td>Mammal</td>
<td>Adult</td>
<td>Testis</td>
<td></td>
</tr>
<tr>
<td>Rat (Wistar)</td>
<td>Mammal</td>
<td>Adult</td>
<td>Testis</td>
<td></td>
</tr>
<tr>
<td>Pig (Sos scrofa)</td>
<td>Mammal</td>
<td>Juvenile</td>
<td>Testis</td>
<td>Testis - A. Archibald, AFRC Roslin, Midlothian</td>
</tr>
<tr>
<td>Chick (Gallus domesticus)</td>
<td>Bird</td>
<td>8-14 day Embryo</td>
<td>Mesonephros /</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Metanephros and Gonad combined</td>
<td></td>
</tr>
<tr>
<td>Alligator (Alligator mississippiensis)</td>
<td>Reptile</td>
<td>Stage 20 Embryo</td>
<td>Gonad</td>
<td>RNA - AM. Coriat, University of Manchester</td>
</tr>
<tr>
<td>Frog (Xenopus laevis)</td>
<td>Amphibian</td>
<td>Stage 45 Tadpole</td>
<td>Whole tadpole</td>
<td>Eggs - P.M. Gaze, University of Edinburgh</td>
</tr>
<tr>
<td>Frog (Xenopus laevis)</td>
<td>Amphibian</td>
<td>Stage 53 Tadpole</td>
<td>Mesonephros</td>
<td>&quot; &quot; &quot;</td>
</tr>
<tr>
<td>Zebrafish (Brachydanio rerio)</td>
<td>Fish Teleost</td>
<td>Larva</td>
<td>Whole larva</td>
<td>RNA - P. Ingham, ICRF, Oxford Reverse transcribed RNA - T. Jowett, University of Newcastle</td>
</tr>
<tr>
<td>Ascidian (Herdmania momus)</td>
<td>Tunicate</td>
<td>Tadpole</td>
<td>Whole tadpole</td>
<td>RNA - M. Levine, Queensland Institute of Medical Research, Australia</td>
</tr>
</tbody>
</table>
way of isolating multiple sequences and therefore for producing sequence comparisons. The zinc finger motif is one of the most highly conserved regions within WTI when human and mouse are compared and therefore is most likely to hybridise cross species. As mentioned in chapter 3, the numbers of this motif are high in the vertebrate genome (Crossley and Little 1991). Therefore if reduced stringency screening with the motif was used, non WTI orthologs especially the EGR family (the most similar fingers to WTI) could be isolated. The use of two primers in PCR allows for the selection of zinc fingers at one site and WTI specificity in the other primer.

4.1.2 Experimental design

The main aim of the design was to favour the amplification of WTI-like genes over any other genes in PCR i.e. to find the balance between specificity and degeneracy. If non specific amplification occurs, in addition to obscuring the desired product, it can inhibit its amplification by sequestering substrates.

4.1.2.1 Template

Three PCR strategies were considered. PCR from genomic DNA amplifying across introns (interexonic); amplifying each exon separately (intraexonic); and RT-PCR using RNA from a tissue likely to be expressing WTI. RNA was chosen as the starting material rather than genomic DNA. This was because the genomic structure of only the zinc finger region was known (Haber et al., 1991). For both interexonic and intraexonic PCR a knowledge of the genomic structure is needed.

Total RNA was prepared wherever possible from tissue found to express WTI in mammals (Table 4.1). The gonads were frequently chosen as these are easily recognised and WTI expression had been demonstrated to be at a higher level than in the kidney of the post natal mouse. The species chosen represent every gnathostome class and one non-vertebrate chordate. Mammalian species were included mainly to act as a test system to show that cross species PCR of WTI was possible. Other species were chosen as the most studied example of their class and are therefore widely obtainable and the most useful for subsequent studies.

4.1.2.2 Primer design

Probably the most important feature of cross species PCR is the design of the oligonucleotide primers. It has been found that for successful priming mismatches are not tolerated in the three 3' bases for a 20bp primer (Sommer and Tautz 1989, Sarker et al., 1990), although some priming can sometimes be obtained in longer primers
Figure 4.1 Predicted amino acid sequence of human WTI showing the exon structure and the positions of degenerate PCR primers used in this study

Exon 1
1 MGSDVRDLNA LLPAVPSLGG GGGCALPVSG AAQWAPVLDF APPGASAYGS
B312→
51 LGGPAPPFPAP PPPPPPPPPHS FIKQEPSWGG AEPHEEQCLS APTVHFSQGF
101 TGTAGACRYG PFGPPPPPSQA SSGQARMFPN APYLPSCLES QPAIRNQGYS

Exon 2
Exon 3
151 TVTFDGTPSY GHTPSHAAQ FPNHSFKHED PMGQQGSLGE QQYSVPPPVY

Exon 4
B300→
201 GCHTPTDSCGGSQLLLRTP YSSDNLYQMT SQLECMHTNQ MNLGATLKGV

Exon 5
Exon 6
251 AAGSSSSSVKW TEGQSNHSTG YESDNHTTPI LCQAZYRIHT HOVFRGQDV

Exon 7
Exon 8
Exon 9
301 RRVPGVAPTL VRSASETSEK RPFMCAYPGC NKRYFKLSHL QMHRSKHTGE

Exon 10
351 KPYQCDFKDC ERRFSRSDLQ KRHQRRHTGV KPPQCKTCQR KPSRSIDHLKT

Exon 1
Exon 2
380 HTRHTGTGTS EKPFSRCPWS CQKKFARSDE LRHHNMHQR NMTKLQLAL*
with a mismatch in the third base from the 3' end. Mismatches between primer and template in any position will affect the melting temperature of the primer template hybrid. In the 5' region mismatches are tolerated as shown by the inclusion of restriction enzyme recognition sites (Scharf et al., 1986).

For WTI, primers were designed by comparison of the two known sequences, mouse (Buckler et al., 1991) and human (Call et al., 1990, Gessler et al., 1990) using both the nucleotide and predicted amino acid sequences. Figure 4.1 shows the positions of the chosen primers relative to human WTI. To minimise the number of mismatches regions of greatest conservation were used and sequences selected containing amino acids with the least number of codon degeneracies. The downstream primer was made complementary to a region within the zinc fingers in the linker region between ZF3 and 4 just downstream of the KTS alternative splice. This linker is a highly conserved motif between the TFIIIA-like zinc fingers (Choo and Klug 1993). It was hoped that the proximity to a splice site might lead to greater conservation at the nucleotide level. The positioning of the upstream primers was not as obvious. Therefore two primers were made complementary to parts of the putative leucine zipper, which is methionine rich (methionine is encoded by only one codon). Further upstream the sequence is GC and proline rich. The former is known to inhibit PCR of WTI (W. Bickmore personal communication); the latter increases the possible degeneracy of the coding sequence and therefore increases the likelihood of mismatches and failure of amplification. One primer was made complementary to the sequence thought least likely to be degenerate in this region. When variation in nucleotide sequence could be predicted from codon degeneracy, or there were changes between human and mouse sequences, mixed bases or inosine were incorporated (Lee et al., 1988, Sommer and Tautz 1989). Inosine is capable of complementing all four bases. Codon preference can sometimes be useful in defining whether codons are likely or not to appear in a species. For those vertebrate species used there only seems to be slight differences in codon preference. No codons were exclusively favoured or unused (Wada et al., 1991). This is different to the situation in some organisms such as E.coli or yeast where some codons are used exclusively and some not at all, arising from differences in proportions of tRNA molecules for the different codons (Ikemura 1985). In humans the GC content of the genome is known to influence codon choice. The primers were designed to amplify all of the species selected rather than targeting one. At the same time primer degeneracy was kept to a minimum as the number of sequences to which primers can anneal increases very rapidly with increasing degeneracy. To produce specific amplification the right
Table 4.2 Sequence of degenerate primers for cross species PCR

<table>
<thead>
<tr>
<th>Primer Number</th>
<th>Position in human sequence</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>B312</td>
<td>Exon 1 (572-599bp)</td>
<td>T\5 GCCGCTGACCCCCACCTCITTCATCAAACA 3'</td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>CGCGGCCGCCTCCTCCTTCTCATCAAACA</td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>CGCGGCCACCCCCACTCCTTCTCATCAAACA</td>
</tr>
<tr>
<td></td>
<td>aa sequence</td>
<td>PPPHSFIKQ</td>
</tr>
<tr>
<td>B300</td>
<td>Exon 4 (1042-1065bp)</td>
<td>G\5 AGCGTCGACAATTTATACACCAATG 3'</td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>AGCACTGACATTTATACACCAATG</td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>AGCACTGACATTTATACACCAATG</td>
</tr>
<tr>
<td></td>
<td>aa sequence</td>
<td>SSDNLQ</td>
</tr>
<tr>
<td>B297</td>
<td>Exon 4 (1075-1101bp)</td>
<td>C\5 TGGCTGGACATGACCTGGAATCAGATG 3'</td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>CTTGAATGACATGACCTGGAATCAGATG</td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>CTTGAATGACATGACCTGGAATCAGATG</td>
</tr>
<tr>
<td></td>
<td>aa sequence</td>
<td>LECMWNQM</td>
</tr>
<tr>
<td>B298</td>
<td>Exon 10/ZF4 (1609-1634bp)</td>
<td>C\5 TGGATGAGCTTCAAGCTGAAAGGTTTTTC 3'</td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>GGCACCCGGACAGCTGAAGGTGTTTTC</td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>TGCCACCCGGACAGCTGAAGGTGTTTTC</td>
</tr>
<tr>
<td></td>
<td>aa sequence</td>
<td>PWRCSPFKE</td>
</tr>
</tbody>
</table>

Underlined sequences indicate changes made to produce restriction enzyme recognition sites
balance between selectivity and degeneracy must be found. To facilitate cloning SalI restriction enzyme recognition sites were designed three bases in from the 5' end of the upstream primers and a HindIII site in the downstream primer. Table 4.2 shows the primers and the sequence from which they were designed.

4.1.2.3 Reaction conditions

An estimate of the melting temperature (Tm) of a primer template hybrid can be calculated on the basis of the primer sequence (Sambrook et al., 1989). This can be used as a guideline to define the annealing temperature of the PCR. With degenerate primers only an average value can be obtained as more than one the primer sequence is present. The Tm is affected by mismatches; the magnitude of the effect can depend on the position of the mismatch within the primer. Therefore conditions at which primers produced specific amplification of WTI were obtained empirically. The stringency of the reaction could be altered by changing either the annealing temperature or the magnesium concentration, both of which influence the binding of primer to the template.

4.1.3 Testing PCR primers

All three upstream primers (B312, B300, B297) were tested in combination with the downstream primer (B298) using either a mouse Wti cDNA clone (pK/S2) and/or total RNA reverse transcribed by the first strand synthesis method (RT-PCR) (using the protocol described in Innis et al., 1990). A range of conditions were tried the most successful, producing specific amplification of Wti, were 94°C 1 minute, 49°C 2 minutes, 72°C 1 minute for 30 cycles and a final extension of 20 minutes at 72°C (in a magnesium concentration of 1.5mM). 'Hot start' PCR where the enzyme is not added until after the first denaturation step was also used. This is in order to avoid non specific extension by the enzyme that can occur at room temperature (Arnheim and Erlich 1992). That Wti was represented within the PCR products was verified by Southern blotting and hybridisation to the mouse Wti cDNA clone (Figure 4.2).

4.1.4 Isolation of rat, pig and alligator WTI orthologs

4.1.4.1 Cross species PCR

Using the conditions which were found to amplify mouse Wti, RT-PCR was performed on rat, pig, chick, alligator and Xenopus RNA. Unfortunately no specific
amplification products were detected. RT-PCR was thought not to be as efficient as conventional PCR from double stranded DNA (I. Jackson personal communication). Therefore cDNA was produced from total RNA using a Boehringer Mannheim cDNA synthesis kit. After 30 cycles of PCR, bands on an agarose gel of the predicted size could be detected with the combination of primers B297-B298 with rat, pig and alligator cDNA templates and with B312-B298 in rat (Figure 4.3a shows a typical amplification with strong specific bands being produced). The possibility of that these PCR products were homologous to WTI was confirmed initially by Southern blotting and reduced stringency hybridisation with the probe Wtl. That non specific amplification was also occurring was evident from the appearance of other non hybridising bands. No specific amplification was detectable using chick or Xenopus cDNA as a template either by ethidium bromide staining or by hybridisation to Wtl. The same negative result was obtained with all combinations of primers, using a range of temperatures, primer and template concentrations.

4.1.4.2 Cloning PCR products

For all successful amplifications, specific WTI products were purified away from non-specific products, primers, nucleotides and Taq polymerase by excision of the band from an agarose gel followed by purification away from the agarose using GENE CLEAN (Bio 101). If the amount of isolated DNA was insufficient for subsequent manipulations a second round of PCR was performed using the purified fragment. Products were digested with SalI and HindIII restriction enzymes to cut within the sites included in the ends of the PCR primers and then ligated into the appropriately digested and dephosphorylated pBluescript SKII+ vector. On transformation of JM101 or XL1-Blue cells no clones were obtained. The exact reason for this failure is not known, but it could be due to the enzymes not cutting these sites near the end of DNA primers. However both HindIII and SalI sites have been used successfully in this context before (Lee et al., 1988 and personal experience). The second possibility became clear when these products were eventually cloned by another method. It was found that very few of the primers were full length, in particular the 5' end, which includes the restriction site, was often lost. If this was also the case in the uncloned product it would mean that many fewer product molecules would be substrates for the restriction enzymes. Therefore fewer molecules with compatible sticky ends would be generated, reducing cloning efficiency.

Blunt ended cloning following the 'double GENE CLEAN' protocol (Bio101) was subsequently used. The ends of PCR products were chewed back and kinased using
Figure 4.3 Isolation of WTI orthologs

a Ethidium bromide stained agarose gel of PCR products amplified from cDNA from mouse, rat, chick, alligator and *Xenopus*, * indicates the presence of the expected size band on the gel. No amplification with these primers was obtained with *Xenopus* cDNA, and with chick cDNA only a smear of non-specific amplification was obtained.

b Reduced stringency hybridisation with *WTI (KS/2)* to white colonies after transformation with rat PCR products ligated into *pBluescript SKII+*. Filters were hybridised and washed at 61°C and washed to 0.5xSSC (2hr exposure).

c PCR products resolved on an agarose gel. PCR analysis of the bacterial colonies giving a positive signal after hybridisation to *WTI*. The forward and reverse M13 vector primers were used to amplify across the cloning site. The expected sizes of the products, depending on the PCR product cloned, are indicated by arrows. Most of the colonies were found to contain inserts of the expected size.
The ragged ends of the purified PCR products were chewed back and the 5' ends kinased before ligation into EcoRV cut pBluescriptSKII+. After transformation white colonies were selected and screened for WTI inserts by low stringency hybridisation. The inserts from positive colonies were analysed by PCR across the cloning site.

**Figure 4.4 Blunt ended cloning strategy used for cloning PCR products**

- PCR Product
  - chew back and kinase
  - Primer
  - 5' → 3'
  - Primer
  - EcoRV Digested pBluescript SKII+
  - +
  - 5' → 3'
  - Blunt End ligation
  - PCR clone
  - Religated vector
  - Eco RV
  - Transform into XL1 Blues
    - Antibiotic and blue/white selection
    - Screen white colonies by low stringency hybridisation
    - Analyse hybridising colonies by minipreps and restriction digest or PCR using vector primers
DNA polymerase I and polynucleotide kinase prior to ligation with EcoRV digested pBluescript SKII+ (Figure 4.4). Chewing back the DNA ends was performed because Taq polymerase, in common with several other polymerases can add a non template directed nucleotide (normally dATP) to the 3' OH of blunt ended double stranded DNA (Clarke 1988). This would interfere with blunt ended ligation. After transformation, recombinants were identified using the blue/white system which relies on the α complementation of β-galactosidase function. Recombinants were screened for WTI-like inserts by reduced stringency hybridisation to Wti (Figure 4.3b). Plasmid DNA from positive clones was prepared by small scale plasmid preparations. The presence of an insert was then confirmed by restriction digest, or PCR across the cloning site using vector primers (Figure 4.3c). Inserts of the expected size were then sequenced.

All the clones sequenced from the three species (rat, pig, and alligator), contained inserts which encoded predicted open reading frames. These closely resembled but differed from the two known primary amino acid sequences of WTI and also differed from each other.¹ Table 4.3 shows the homology at the nucleotide and amino acid levels. These were named RWTI, (rat WTI), PWTI (pig WTI) and AWTI (alligator WTI). This high level of similarity at the nucleotide and amino acid level provided strong evidence that orthologs of WTI had been isolated. More than one clone was sequenced for each species as Taq polymerase has a higher rate of misincorporation as it lacks 3'5' exonuclease (proof-reading) activity. The error rate is currently estimated to be 8.5x10⁶ but is known to vary with the conditions (Fuchareon et al., 1989). Therefore for a 200bp fragment an estimated 3% of cloned products will contain errors. This error rate was consistent with the very low level of variation between clones e.g. in alligator only one point of variation in three 500bp clones was found. On publication the sequence of the entire coding region of WTI from a rat cDNA (Sharma et al., 1992) corresponded to that obtained here. Sequencing several clones also revealed variation (Figure 4.5) owing to the presence of the two alternative splices which have already been observed in mouse and human (Haber et al., 1991).

¹The sequences of all the clones and a nucleotide comparison are contained in Appendix A.
Table 4.3 Similarity of the PCR clones isolated to human and mouse *WT1* at the nucleotide and amino acid level.
Figures were calculated using the Gap program from GCG.

<table>
<thead>
<tr>
<th>Species</th>
<th>Nucleotide % Identity</th>
<th>No. Amino Acids compared</th>
<th>Amino Acid % Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Human: 91.5</td>
<td>Mouse: 95.2</td>
<td>Human: 99.1</td>
</tr>
<tr>
<td>Rat</td>
<td>91.5</td>
<td>95.2</td>
<td>335</td>
</tr>
<tr>
<td>Pig</td>
<td>95.1</td>
<td>91.9</td>
<td>165</td>
</tr>
<tr>
<td>Alligator</td>
<td>81.0</td>
<td>81.5</td>
<td>156</td>
</tr>
<tr>
<td>Chick</td>
<td>82.5</td>
<td>80.3</td>
<td>154</td>
</tr>
<tr>
<td><em>Xenopus</em></td>
<td>84.2</td>
<td>83.8</td>
<td>82</td>
</tr>
<tr>
<td>Axolotl</td>
<td>85.2</td>
<td>84.7</td>
<td>72</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>82.7</td>
<td>82.2</td>
<td>72</td>
</tr>
<tr>
<td>(predicted cDNA sequence)</td>
<td>82.7</td>
<td>82.2</td>
<td>72</td>
</tr>
</tbody>
</table>
Table 4.4 PCR primers designed using sequence from PCR clones

<table>
<thead>
<tr>
<th>Primer Number</th>
<th>Position in human sequence</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>C227</td>
<td>Exon 6 (1221-1244bp)</td>
<td>5'CTGCGGAGCCCAATACAGAATACA 3'</td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>CTGCGGAGCCCAATACAGAATACA</td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>CTGTTGTCGCTGATACAGAATACA</td>
</tr>
<tr>
<td></td>
<td>Alligator</td>
<td>CTGTTGACCCGTTACACAGAATACA</td>
</tr>
<tr>
<td></td>
<td>aa sequence</td>
<td>CGAQYRIH</td>
</tr>
<tr>
<td>B620</td>
<td>Exon 7 / ZF1 (1324-1352bp)</td>
<td>5'GAG_CAGCTGAGAAACGCCTTCATGTG 3'</td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>GAGACCAGTGAGAAACGCCTTCATGTG</td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GAACCAGTGAGAAACGTCTTCATGTG</td>
</tr>
<tr>
<td></td>
<td>Alligator</td>
<td>GAGCAAATGAAAAACGTCTTCATGTG</td>
</tr>
<tr>
<td></td>
<td>aa sequence</td>
<td>ETSEKRPMC</td>
</tr>
<tr>
<td>C229</td>
<td>Exon 7 / ZF1 (1384-1406bp)</td>
<td>5'AAGCTCTCCCAATTACAGATGCA 3'</td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>AAGCTCTCCCAATTACAGATGCA</td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>AAGCTCTCCCAATTACAGATGCA</td>
</tr>
<tr>
<td></td>
<td>Alligator</td>
<td>AAGTTATCCTCCATTTACAGATGCA</td>
</tr>
<tr>
<td></td>
<td>aa sequence</td>
<td>KLISHQMH</td>
</tr>
<tr>
<td>C228</td>
<td>Exon 9 / ZF3 (antisense)  (1579-1599bp)</td>
<td>5'ACCTGTATGAGTCTCTGGTG 3'</td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>ACCTGTATGAGTCTCTGGTG</td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>ACCTGTATGAGTCTCTGGTG</td>
</tr>
<tr>
<td></td>
<td>Alligator</td>
<td>ACCTGTATGAGTCTCTGGTG</td>
</tr>
<tr>
<td></td>
<td>aa sequence</td>
<td>GTHTRTH</td>
</tr>
</tbody>
</table>

Underlined sequences indicate changes made to produce restriction enzyme recognition sites.
Bold letters indicate zinc chelating residues within zinc fingers.
Figure 4.5 Schematic representation of the human \textit{WTI} cDNA and the \textit{WTI} clones isolated by degenerate PCR.

The primers with which successful amplifications were obtained are shown as well as the alternative splice forms observed. The clones which were subsequently used as probes are also indicated (*).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{schematic.png}
\caption{Schematic representation of the human \textit{WTI} cDNA and the \textit{WTI} clones isolated by degenerate PCR.}
\end{figure}

### PCR Primers

<table>
<thead>
<tr>
<th>PCR Primers</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' 3'</td>
<td></td>
</tr>
<tr>
<td>B312 B298</td>
<td>Rat</td>
</tr>
<tr>
<td>B297 B298</td>
<td></td>
</tr>
<tr>
<td>B297 B298</td>
<td>Pig</td>
</tr>
<tr>
<td>B300 C228*</td>
<td>Chick</td>
</tr>
<tr>
<td>C227 C228</td>
<td></td>
</tr>
<tr>
<td>C229 C228</td>
<td></td>
</tr>
<tr>
<td>B297 B298*</td>
<td>Alligator</td>
</tr>
<tr>
<td>B620 B298*</td>
<td>Xenopus</td>
</tr>
<tr>
<td>C229 C228</td>
<td></td>
</tr>
<tr>
<td>B620 C228</td>
<td>Zebrafish</td>
</tr>
<tr>
<td>C229 C228*</td>
<td></td>
</tr>
</tbody>
</table>

### Alternative Splices

<table>
<thead>
<tr>
<th>17aa</th>
<th>KTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>N/D</td>
</tr>
<tr>
<td>-</td>
<td>N/A</td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>+</td>
<td>N/A</td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>N/A</td>
<td>N/D</td>
</tr>
</tbody>
</table>

N/A = region not covered by clone
N/D = not observed but not large sample examined
* Probes used for library screens
+ Not to scale
4.1.5 Redesign of PCR primers

The extra sequence information provided by these different species, in particular the first non mammalian *WTI* to be isolated, *AWTI*, enabled several more primers to be designed from conserved regions. Upstream primers C227, B620 and C229 were synthesised, the latter two were complementary to ZF1. A second downstream primer C228 with the 3' end within one of the invariant zinc chelating histidines in ZF3 was also made (Table 4.4, Figure 4.1). These primers were tested on alligator eDNA and found to give specific amplification of *AWTI* (data not shown).

4.1.6 Isolation of chick, *Xenopus* and axolotl *WTI* orthologs

Amplification using chick cDNA from 8-14 day embryos was successful with all the combinations of new primers and the combination B300-C228. That B300-B298 combination failed to work but B300-C228 gave a strong specific signal suggests that there was a mismatch between the chick sequence and B298. Amplification with *Xenopus* cDNA was successful with only the primer combinations B620-B298 and C229-C228. Two sources of total RNA were used; either whole stage 45 tadpoles or mesonephros from stage 53 tadpoles (Nieuwkoop and Faber 1967). Only RNA from the mesonephros produced specific amplification, suggesting that in the whole tadpole RNA the concentration of template was too low. Figure 4.6 shows *WTI* specific amplification from chick and *Xenopus* cDNA. These PCR products were cloned and sequenced as above. Again sequences resembling but not identical to previously cloned *WTI* genes were obtained (Table 4.3, Appendix A). These were named *CWTI* and *XWTI* for chick and *Xenopus* respectively. In both cases only the shortest transcript was isolated (Figure 4.5) even though the 17aa region was covered in chick and the KTS region in *Xenopus*.

These primers, along with *XWTI*, were supplied to P.A. Tsonis at the University of Dayton, Ohio who subsequently amplified and cloned 216bp of *WTI* from the axolotl (*Ambystoma mexicanum*), *AXWTI* (Table 4.3).

4.1.7 Isolation of the zebrafish *WTI* ortholog

4.1.7.1 Degenerate PCR

The presence of *WTI* in the zebrafish, and the ascidian (a non-vertebrate chordate) was investigated by PCR from whole larva or tadpole cDNA. In both cases
Figure 4.6 Isolation of chick and *Xenopus* WTI orthologs

Resolution of PCR products on an agarose gel. PCR used the primers designed on the basis of the clones already obtained. Strong bands and a low level of non-specific amplification were observed.

<table>
<thead>
<tr>
<th>Lane</th>
<th>PCR Template</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chick cDNA</td>
<td>C229/C228</td>
</tr>
<tr>
<td>2</td>
<td>Chick cDNA</td>
<td>B300/C228</td>
</tr>
<tr>
<td>3</td>
<td><em>Xenopus</em> mesonephric cDNA</td>
<td>B620/B298</td>
</tr>
<tr>
<td>4</td>
<td><em>Xenopus</em> mesonephric cDNA</td>
<td>C229/C228</td>
</tr>
<tr>
<td>5</td>
<td>Chick cDNA (RePCR)</td>
<td>C229/C228</td>
</tr>
<tr>
<td>6</td>
<td>Alligator cDNA</td>
<td>C229/C228</td>
</tr>
<tr>
<td>7</td>
<td>No DNA (negative) control</td>
<td>C229/C228</td>
</tr>
</tbody>
</table>

M 1 2 3 4 5 6 7

kb

1.35

1.08

0.87

0.60

0.31

0.28

0.23

0.19
amplification was obtained, but on cloning these products either non WTI-like products or contaminants were found.

To avoid contamination problems, PCR on zebrafish was repeated at the source of zebrafish material in Newcastle. A range of conditions and primer combinations which had been known to produce WTI specific amplification were used with either whole embryo RNA (reverse transcribed by first strand synthesis of RACE reaction) or genomic DNA (both a gift from T. Jowett). Genomic DNA was used in case PCR from whole embryo RNA was not successful. The primers used were known to cross introns in the human genomic structure therefore PCR products would probably not be the same size as any contaminants, but this meant the size of the WTI specific products could not be predicted.

Very little amplification with the RNA derived template was seen even with nested PCR (Mullis and Faloona 1987). In contrast a wide range of bands were observed with genomic DNA (Figure 4.7a). To identify which of these bands might be WTI orthologs, Southern blotting and reduced stringency hybridisation to AWT1 was performed. By comparison of the ethidium stained agarose gel and the autoradiograph two genomic bands were identified that hybridised to AWT1 far more strongly than any others (Figure 4.7b). These corresponded to amplifications with primers B620-C228 and C229-C228. These primers were ones that had given amplification with Xenopus WTI (the most diverged WTI gene so far isolated). Both bands were very faint, this could well have been due to the large size, approximately 2.8kb, of the products. PCR is known to favour the amplification of smaller products as extension as far as the other primer is more likely to occur for shorter regions. The similar size of the two products was consistent with the human exon structure, in which B620 and C229 would be in the same exon separated by 60bp (Figure 4.9a).

4.1.7.2 Turbo cloning

The B620-C228 band was purified away from non-specific products as before (section 4.1.4.2). This was used as template in heminested PCR (Li et al., 1990) with primers C229-C228 to produce sufficient material for cloning. It was found for successful rePCR of such a long product the fragment could not be directly visualised with ethidium bromide and UV light, as nicking of the DNA inhibited the second PCR resulting in a smear, or a ladder of smaller bands.

Most conventional cloning techniques can be size selective because the ends of smaller fragments have a greater chance of being ligated to the ends of the same vector DNA molecule. To avoid this turbo cloning was used (Boyd 1993) (Figure
Figure 4.7 Isolation of zebrafish WT1

a Agarose gel electrophoresis of zebrafish PCR products. RT-PCR and PCR from genomic DNA were tried using two annealing temperatures and a range of different primer combinations. Multiple bands were obtained with genomic DNA but none with RT-PCR after 30 rounds of amplification.

b Identification of WT1 specific bands. Autoradiograph of PCR products after hybridisation with AWT1 at low stringency. Digestions of WT1 clones were also run on the gel for comparison of the level of hybridisation to the probe (1.5hr exposure).

→ identifies two bands showing strong hybridisation to AWT1, but were only just visible on the ethidium bromide stained gel above.
Figure 4.8 Turbo cloning of PCR products

The ragged ends of the PCR products were chewed back and the 5' ends kinased. These were ligated to SmaI digested pBS:lox in conditions of macromolecular crowding. To the ligation products Cre enzyme was added which catalysed recombination of the lox sites leading to circularisation of molecules where the lox sites were in the same orientation. After transformation, white colonies were screened for WTI inserts by low stringency hybridisation to AWTI and positive colonies were further analysed by restriction digest of the plasmid DNA.
4.8). This involves blunt ended ligation in conditions of macromolecular crowding into pBS:lox, where the ends of the insert can rapidly independently ligate to vector molecules. Circularisation is achieved afterwards by utilisation of the Cre:lox recombination system of P1 bacteriophage. Cre enzyme is added and recognises lox sites in the vector, and those with lox sites in the same orientation are circularised. After electroporation into XL1-Blue cells recombinants were screened for inserts with the blue/white selection system and for WTI specific inserts by hybridisation to AWTI at low stringency. About 10% of white colonies were found to contain inserts that hybridised to the probe. These were analysed by purification of the plasmid by 'Magic Minipreps' (Promega) and restriction digest. 50% of these had the same restriction pattern and the inserts were the expected size. Sequencing of the inserts using vector primers showed the presence of the PCR primers (C228/C229) at either end of the insert, confirming these as full length products.

From comparison with the human genomic structure it was expected that the clones would contain sequences homologous to the three exons 7, 8 and 9 encoding ZF1-3. Sequencing of the genomic clones revealed three regions of homology corresponding to these exons. These, as expected, were separated by unconserved sequences when compared with the human genomic sequence (Gessler et al., 1992). The AT richness and repetitive nature of the sequences suggested that they were indeed introns. The regions of homology were found to extend just outside the exons to include the splice donor and acceptor sequences. From this it was predicted that the zebrafish WTI cDNA had a similar exon structure to human, although the actual sizes of the introns were completely diverged (Figure 4.9). The 60bp between primers B620 and C229 was amplified by heminested PCR using B620 and a ZF2 primer designed from the sequence of the cloned region. The product was sequenced directly (Winship et al., 1989) and found to correspond to the part of ZF1 expected.

4.2 Isolation by library screening

4.2.1 Introduction

PCR is a fast and relatively easy method for obtaining conserved regions of a gene. From the sequence data obtained above it was clear that WTI has not evolved quickly within the vertebrates, especially within the zinc fingers (Table 4.3). This is consistent with mutation analysis of WTI, where almost all mutations found in WTI affect the zinc fingers. To isolate more of the coding sequence, to determine the extent and pattern of conservation within the proposed transregulatory domain,
Figure 4.9
Comparison of human and zebrafish *WTI* genomic structures in the zinc finger region

a Human genomic structure showing positions of primers producing *WTI* specific amplification in zebrafish.

```
<table>
<thead>
<tr>
<th>Human genomic structure showing positions of primers producing <em>WTI</em> specific amplification in zebrafish.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exon 7</strong></td>
</tr>
<tr>
<td><strong>ZF1</strong></td>
</tr>
<tr>
<td><strong>Human</strong></td>
</tr>
<tr>
<td><strong>PCR primers</strong></td>
</tr>
</tbody>
</table>
```

b Genomic structure of zebrafish *WTI*

```
<table>
<thead>
<tr>
<th>Zebrafish genomic structure showing positions of primers producing <em>WTI</em> specific amplification in zebrafish.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ZF1</strong></td>
</tr>
<tr>
<td><strong>80bp</strong></td>
</tr>
<tr>
<td><strong>PCR clone</strong></td>
</tr>
<tr>
<td><strong>pBS:lox</strong></td>
</tr>
<tr>
<td><strong>SaiI</strong></td>
</tr>
</tbody>
</table>
```

Comparison with the human intron/exon structure shows conservation of the organisation of the exons but not the intron sizes in the PCR clone *ZWT1*. The SaiI/RsaI fragment used to probe the cDNA libraries is shown.
several strategies could be used. Library screening was chosen in preference to PCR based techniques such as 5' RACE (Rapid amplification of cDNA ends) because of the difficulty experienced amplifying this GC rich region in humans and the possibility of contamination. The major disadvantage to cross species isolation by library screening had been removed by the isolation of homologous probes.

4.2.2 Library screens

cDNA libraries in bacteriophage vectors were obtained either from the embryonic stages most likely to be expressing WTI mRNA or from adult testis (Table 4.5). These libraries were titred on the relevant host cells, and approximately $4 \times 10^5$ clones screened following the recommended method supplied by Stratagene for λZAPII. Cloned PCR products were used to probe the library of the same species under high stringency conditions. Clones giving positive signals were purified with further rounds of screening. On isolation the phage were amplified and either converted into phagemids by in vivo excision (λZAP type vectors) or the inserts were subcloned into pBluescript SKII+ at the EcoRI site (λgt11 type vectors). The isolates were then sequenced. The library from the striped-faced dunnart (*Sminthopsis macroura*), a marsupial that resembles a mouse, was initially screened by S. Wilcox at La Trobe University, Australia with the human WTI probe (WT33) at reduced stringency. Five positives from this primary screen were then analysed for WTI by PCR directly from the bacteriophage and a single clone isolated by hybridisation at reduced stringency to AWTI.

4.2.3 WTI cDNA clones isolated

The results of the library screens are summarised in Table 4.5 and Figure 4.10. The percentage similarities to mouse and human WTI are shown in Table 4.6, calculated using the gap program of GCG\(^1\).

A unique clone was isolated from the chick (C2.1), the alligator (AL1), and the marsupial mouse (Sc41). The three partial cDNA clones between them cover the entire coding region, the entire 3' non coding region and probably most of the 5' non coding region, with large amounts of overlap between them. The alligator and chick clones were found to contain the regions isolated by PCR and their sequence was found to be identical to the consensus sequences obtained from multiple PCR clones.

\(^1\)All sequences are contained in Appendix A
Table 4.5 Summary of libraries screened for WTI and the clones isolated

<table>
<thead>
<tr>
<th>Species</th>
<th>Taxon</th>
<th>Vector for Library construction</th>
<th>Source of cDNA</th>
<th>Clone Name / Size</th>
<th>Isolated Region covered</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alligator (Alligator mississippiensis)</td>
<td>Reptile</td>
<td>λ ZAP II (Stratagene custom made)</td>
<td>Embryonic - Stage 20</td>
<td>AL1 / 1.3kb</td>
<td>Exon 2 to 3' untranslated region</td>
<td>Paul Sharpe University of Manchester</td>
</tr>
<tr>
<td>Chick (Gallus domesticus)</td>
<td>Bird</td>
<td>λgt11</td>
<td>Embryonic - Stage 20</td>
<td>C2.1 / 1.6kb</td>
<td>5' untranslated region to exon 9</td>
<td>Nick Platt University of Oxford</td>
</tr>
<tr>
<td>Frog (Xenopus laevis)</td>
<td>Amphibian</td>
<td>λgt10</td>
<td>Neuralation - Stage 17</td>
<td>None</td>
<td>None</td>
<td>Doug Melton Harvard University, USA</td>
</tr>
<tr>
<td>Marsupial Mouse (Sminthopsis macroura)</td>
<td>Marsupial</td>
<td>λ ZAP</td>
<td>Testis</td>
<td>Sc41 / 2kb</td>
<td>Exon 3 to poly A tail</td>
<td>Stephen Wilcox, Jenny Marshall Graves, La Trobe University, Australia</td>
</tr>
<tr>
<td>Zebrafish (Brachydanio rerio)</td>
<td>Fish</td>
<td>λ ZAP II</td>
<td>Adult Post somitogenesis</td>
<td>Z1a / 3.5kb</td>
<td>Genomic Intron 6 to 9</td>
<td>David Grunwald University of Utah, USA</td>
</tr>
</tbody>
</table>
Figure 4.10 Diagram of WTI clones isolated by library screening, in relation to the human WTI gene.

The presence of alternatively spliced sequences within the clones is indicated.
Table 4.6 Similarity of clones isolated from libraries to human and mouse *WTI* at the nucleotide and amino acid level.

Figures were calculated using the Gap program from GCG.

<table>
<thead>
<tr>
<th>Species</th>
<th>Nucleotide % Identity</th>
<th>No. Amino Acids Compared</th>
<th>Amino Acid % Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Human</td>
<td>Mouse</td>
<td>Human</td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td>84.8</td>
<td>84.8</td>
<td>240</td>
</tr>
<tr>
<td>Chick</td>
<td>83.4</td>
<td>82.5</td>
<td>390</td>
</tr>
<tr>
<td>Alligator</td>
<td>84.0</td>
<td>84.0</td>
<td>289</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>80.0</td>
<td>79.8</td>
<td>112</td>
</tr>
</tbody>
</table>

(predicted cDNA sequence)
Although approximately $1 \times 10^6$ clones were screened from both *Xenopus* cDNA libraries and rescreened at reduced stringency with the non finger region of the alligator clone (NcoI/SalI restriction digest fragment of AL1 covering exons 2-6, Figure 4.10) no positives were found.

4.2.4 Analysis of zinc finger 4 in chick

On sequencing the 3' end of C2.1 it became apparent that the 3' 100bp following ZF3 did not correspond to ZF4 as would have been expected. The sequence encoded a termination codon four residues downstream from the end of ZF3 (Figure 4.11a). The sequence did not correspond to any sequences in the GenBank or EMBL databases and so was not part of the bacteriophage vector accidentally included during subcloning.

This and the previous inability to amplify chick *WT1* with a ZF4 primer B298 (Section 4.1.6) led to doubts about the conservation of ZF4 in the chick. To test this another ZF4 primer (D502) was designed using the closely related alligator sequence (Figure 4.11a). PCR with a chick cDNA template showed specific amplification of the expected size band (Figure 4.11b). The PCR product was directly sequenced and was found to encode the 'missing' zinc finger. That this product was derived from chick and no other contaminating species was verified by comparing the sequence of ZF3 from the PCR product. The sequences of the chick PCR clones, C2.1 and the PCR product were found to be identical. On the sequencing gel it was found that once the boundary from ZF4 to ZF3 was crossed two sequences became superimposed on each other (Figure 4.11ci). This was found to be due to the presence of a 9bp insertion encoding the KTS alternative splice in some transcripts (Figure 4.11cii & iii).

The origin of the non ZF4 sequence is not clear but that it commences exactly at the position of a splice site points to it being due to an aberrant splicing event. The possibility that it is due to the intron at this position not being excised was considered. In human the first 9bp of this intron encode the KTS alternative splice. It has been shown that this sequence is also present in chick *WT1* transcripts but no such sequence was present in C2.1. This suggests that the unknown sequence is not this intron. Clone C2.1 could have been produced from contaminating genomic DNA, or from a pseudogene from which ZF4 had been lost. The presence of a pseudogene in the chick genome was investigated. Figure 4.11d shows a zoo blot of EcoRI genomic digests probed at high stringency with *CWT1*. Only one band was detected,
Figure 4.11 Isolation of ZF4 from chick

a Nucleotide and amino acid comparison of C2.1 and human WTI ZF3&4 showing a high level of similarity in ZF3 but not in the region expected to be ZF4.

b PCR amplification of chick ZF4
The three 3' primers were used to try to amplify WTI from chick cDNA in combination with 5' primer C227 (located in exon 6).
In accordance with earlier results amplification was seen with C228 (ZF3) but not B298 (ZF4). In addition amplification of a band the same size as the alligator was seen with the new ZF4 primer D502. This suggested that the unidentified sequence in C2.1 was a cloning or splicing artefact and that ZF4 is conserved in chick.

(D502 5' CATGTTGTGATGACGAACCTAA(C/T)TC(A/G)TC 3')

ci Direct sequencing of the chick cDNA PCR product with D502
A single sequence is seen until the ZF3/4 boundary is crossed, and then the amount of cross banding increases. This was found to be due to two sequences superimposed on each other.

ii&iii Dissection of the two sequences showed that they were both identical to C2.1 in the ZF3 region and that they correspond to the + and - KTS forms of the mRNA, owing to the insertion of the 9bp at the exon 9/10 boundary. The splice site between ZF3&4 was also seen to be the position where the homology between C2.1 and CWT1 ended indicating the probability that the mRNA transcript from which C2.1 was made was aberrently spliced.

<- indicates the positions of the splice sites. * indicates the position of a mismatch between the chick sequence and primer B298 which probably inhibited amplification with this primer.

d PhosphorImager analysis of a zoo blot probed at high stringency with CWT1 (PCR clone). Only a single band was detected indicating that WTI is single copy in the chick genome and that C2.1 could not have been derived from a pseudogene.
(B, bovine, C, chick, A, alligator)
a

**Chick/Human**

GATCATCTGAAGACTCATACCAGGACTCATACAGCTA...TAGTGAAAAGAACATATACATTCAGCTCAACAAATGTCAACTCCACATTCAGTTGTCTTTCTAAAGGCTACCTACGA

**3' Primers**

- C228
- B298
- D502

b

<table>
<thead>
<tr>
<th>Lane</th>
<th>PCR Template</th>
<th>3' Primer</th>
<th>kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chick cDNA</td>
<td>B298</td>
<td>1.35</td>
</tr>
<tr>
<td>2</td>
<td>Chick cDNA</td>
<td>D502</td>
<td>1.08</td>
</tr>
<tr>
<td>3</td>
<td>Chick cDNA</td>
<td>C228</td>
<td>0.87</td>
</tr>
<tr>
<td>4</td>
<td>Alligator cDNA</td>
<td>D502</td>
<td>0.60</td>
</tr>
<tr>
<td>5</td>
<td>Alligator cDNA</td>
<td>C228</td>
<td>0.31</td>
</tr>
<tr>
<td>6</td>
<td>No DNA (negative)</td>
<td>D502</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>No DNA (negative)</td>
<td>C228</td>
<td></td>
</tr>
</tbody>
</table>
Two superimposed sequences

single sequence

++KTS

-ZF3

-ZF4

-ZF3

-ZF4

-WT1
demonstrating the WTI gene to be single copy in the chick genome.

4.2.5 Isolation of a zebrafish WTI genomic clone (Z1a)

To screen the zebrafish libraries a SalI/RsaI fragment from a zebrafish genomic PCR clone was used that contained part of ZF1, ZF2 and the 80bp intron between them (Figure 4.9b). This excluded the long 2.5kb intron between ZF2 and ZF3. Two cDNA libraries were screened, a post-mitogenesis one and an adult one. One positive was obtained; from the adult library. On sequencing the isolated clone (Z1a) it was found to contain interruptions between the zinc finger encoding sequences identical to the predicted introns in the genomic PCR clones ZWT1. This could either be because introns had not been removed due to partial splicing of the pre-mRNA transcript or that the clone was generated from contaminating genomic DNA in the RNA preparation. The presence of four introns within the clone favoured the latter possibility (Figure 4.12a). The clone extends 5' and 3' of the PCR clone containing the whole of exons 7, 8 and 9 and therefore includes the region of intron 9 which encodes the KTS alternative splice in humans. In zebrafish the two alternative splice donor sites are very highly conserved suggesting it is very likely that both + and -KTS isoforms are expressed in fish (Figure 4.12b). Therefore genomic organisation of WTI with each zinc finger in a separate exon has been conserved through vertebrate evolution. In contrast the sizes of the introns are completely different; introns 7 and 8 being 2kb and 600bp respectively in human (Haber et al., 1991) whereas in zebrafish they are 80bp and 2.5kb (Figure 4.9).

4.3 Discussion

PCR and library screening was an effective method for isolating a number of different WTI orthologs. Less sequence was isolated than was desired especially in the transregulatory domain. This was firstly due to PCR in this region not being successful, either because the sequence variation was too great for the primers used, or the GC content was too high. Secondly 5' incomplete clones were obtained from the library screens. This could be a result of the reverse transcription step not producing full length DNA; reverse transcriptase can be inhibited due to the presence GC rich regions and RNA secondary structure. The first clones isolated for mouse and human also lacked part of the 5' region.
Figure 4.12 Genomic structure of zebrafish *WT1*

a Comparison of region covered by the PCR and library clones. Z1a contains 460bp and 200bp of the introns flanking exon 7 (ZF1) and exon 9 (ZF3) and includes the region encoding the KTS alternative splice in human.

![Diagram of genomic structure](image)

b Comparison of the sequence of Z1a with the human genomic sequence showing the conservation of the region containing the KTS alternative splice, indicating that there is a high probability that the inclusion of this alternative splice is conserved across 400Myr.
Much time was wasted because of PCR contamination. Cross species PCR is more prone to contamination than homologous PCR for several reasons.

i) Primers are designed from known and therefore amplified or cloned sequences. These are likely to be more similar to the primers than the desired template and so in the first few rounds of amplification these will be preferentially amplified. This means that even if the template is initially in excess of the contaminant the final product may contain enough contaminant to obscure the real product or take over the reaction.

ii) If only one species is involved the presence of a contaminant can be assayed for by high stringency hybridisation with the suspected contaminant, but where many species are involved it may not be known which is the contaminant.

The precautions advised to avoid contamination were followed (Clackson et al., 1991). It was found that consistent contamination free PCR was only possible by using a separate room for setting up the reactions.

The reason for failure to isolate cDNA sequences from zebrafish and ascidian (tunicate) by PCR may be because of a restriction of high levels of WT1 mRNA to certain structures. In whole embryo preparations WT1 message might be at too low a concentration for cross species PCR. This seems to have been the case for Xenopus as amplification was successful only with RNA isolated from the mesonephros (Section 4.1.6). There are no vertebrate-like kidneys present in the ascidian from which RNA could be isolated. The gonads may be a possible source of material but these gonadal structures may have a completely different origin to vertebrate gonads.

That no larger Xenopus clone was obtained from cDNA library screens was disappointing. This was maybe because WT1 was not expressed in the early stages from which the libraries were made. In the mouse the first Wt1 expression is observed at about 9 days, just before the onset of mesonephric differentiation (Armstrong et al., 1992). In Xenopus the first evidence of pronephros differentiation is seen at stage 23 with the slight thickening of the mesoderm corresponding to the pronephric duct rudiment (Nieuwkoop and Faber 1967, Lynch and Fraser 1990). It was not possible to locate libraries later than stage 24. This is probably owing to the major interest in Xenopus concentrating in the earliest stages of development. The zebrafish postcomitogenesis library was thought likely to contain WT1 clones, although none were detected. It may be that the intron within the probe reduced the affinity of the probe for a cDNA to too large extent. The isolation of the genomic clones has allowed the comparison of WT1 to be extended to genomic structure.
In *Xenopus* there are probably two copies of *WT1* in the genome, arising from a tetraploidisation event which occurred about 30 Myra (Tymowska and Fischberg 1973). In the three *Xenopus* PCR clones which were sequenced three differences were found in one clone relative to the others. It is not known if these arose from PCR errors or if the two forms were isolated. The number of differences was higher than for the three alligator clones sequenced, where one change was noted in clones twice the length of *XWT1*. If these clones do represent the two copies of *WT1* they are very highly conserved. This could be because substitutions within *WT1* zinc fingers are thought to have dominant negative properties, which has lead to selection for conservation.

The sequence information derived here will enable further species to be cloned more easily and also lead to suggestions about the structure and evolution of *WT1*, which are discussed in the next chapter.
Chapter 5

Structural and evolutionary comparison of WT1 orthologs
5.1 Introduction

The sequence information from the \textit{WT1} orthologs obtained can be used to analyse the conservation of structural features of the \textit{WT1} protein. Conservation can then be used to make inferences about the structural and functional importance of these regions. This can be especially useful when combined with functional studies of the gene. The sequences can also be used to investigate the evolutionary changes that have occurred, such as the rate of change of both the nucleotide and the encoded amino acid sequence.

In the following discussion it was found that the most convenient way in which to define the regions referred to was to use the exon structure of the gene. Knowledge of the exon structure is based entirely on the information from the human and zebrafish, as no information has been as yet been obtained for other species. In addition all positions in the amino acid and nucleotide sequence refer to numbering of the human +17aa, +KTS isoform.

5.2 Structural analysis

5.2.1 Sequence motifs

5.2.1.1 Structure prediction

The nucleotide sequences of \textit{WT1} orthologs were translated to produce the predicted amino acid sequence of the protein (Figure 5.1). Most of the amino acid changes fall into the category of frequently accepted, conservative mutations which are due to single base pair mutations when compared against those in Dayhoff's matrix (Dayhoff \textit{et al.}, 1972). Therefore since multiple changes have not occurred, mutations which result in an accepted change must occur infrequently. The sequences were analysed using the motifs program from GCG, but no other sequence motifs were found in addition to the zinc fingers. The hydrophobicity profiles (Kyte and Doolittle 1982) and secondary structure predictions (Chou and Fasman 1978) of each ortholog were compared and no great changes were observed in the different species (Appendix B). An increase in hydrophilicity was seen in the zinc finger region consistent with it interacting with DNA. Protein structure prediction from the primary amino acid sequence alone is still in its infancy. Only a low level of accuracy is obtained with the Chou and Fasman prediction (Creighton 1984) and so a newer and more reliable program was tried. This was Predict Protein (Sander and Schneider 1991) which relies on sequence comparison to predict a structure. The only region
Figure 5.1 Comparison of the predicted amino acid sequences of the WTI orthologs

The human exon structure is shown as are the positions of residues discussed in the text (bold). The alternative splices and known sequence motifs are highlighted.
<table>
<thead>
<tr>
<th>EXON1</th>
<th>MGSDVRDLNALLPAVPSL-GGGGSCALPVSGAAQWAPVLDFAPPASAYGSL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S.G.G.R.</td>
</tr>
<tr>
<td></td>
<td>S.R.</td>
</tr>
<tr>
<td></td>
<td>S.P.NSN.M.S.R.</td>
</tr>
<tr>
<td>polyproline</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>GGPAPPAPPPPPPPPPHSPIKQEPSWGAEPHECQELSAPTVHFGSQPTGT</td>
</tr>
<tr>
<td>Mouse</td>
<td>..G.R.</td>
</tr>
<tr>
<td>Rat</td>
<td>..L.</td>
</tr>
<tr>
<td>Chick</td>
<td>..N.SD.Y</td>
</tr>
<tr>
<td></td>
<td>polyproline</td>
</tr>
<tr>
<td>Human</td>
<td>AGACRYGPFGPPPPSPASSQQARMFPNAPYLPSCLESEQPAIRNGVYSPVVT</td>
</tr>
<tr>
<td>Mouse</td>
<td>T.S.</td>
</tr>
<tr>
<td>Rat</td>
<td>.A.P.P.</td>
</tr>
<tr>
<td>Chick</td>
<td>.N.Q.G.A.</td>
</tr>
<tr>
<td>Alligator</td>
<td>A.A.</td>
</tr>
<tr>
<td>Human</td>
<td>GTPSYGHTPSHAQFNHSFKHEDPMGQQGSLGEOQYSVPPPVYGCHTPTD</td>
</tr>
<tr>
<td>Mouse</td>
<td>.A.</td>
</tr>
<tr>
<td>Rat</td>
<td>.A.</td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td>.Q...S.P.D.</td>
</tr>
<tr>
<td>Chick</td>
<td>.P..A.S...IA.T..D.</td>
</tr>
<tr>
<td>Alligator</td>
<td>EXON4 EXON5</td>
</tr>
<tr>
<td>Human</td>
<td>SCTGSQALLRTPYSSDNLYQMTSOLCEMTWNQMNLAGTKVGAAGSSSVK</td>
</tr>
<tr>
<td>Pig</td>
<td>M.M.</td>
</tr>
<tr>
<td>Mouse</td>
<td>M.M.</td>
</tr>
<tr>
<td>Rat</td>
<td>T..N.</td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td>R..S...</td>
</tr>
<tr>
<td>Chick</td>
<td>N..N.</td>
</tr>
<tr>
<td>Alligator</td>
<td>putative leucine zipper</td>
</tr>
<tr>
<td></td>
<td>EXON6 273 EXON7 281</td>
</tr>
<tr>
<td>Human</td>
<td>WTEGQSNHSTGZEBDNHPTTPIL-CGAQYRIHTHGVFQGMQDOVRVPGVAPT</td>
</tr>
<tr>
<td>Pig</td>
<td>.G...T.A.</td>
</tr>
<tr>
<td>Mouse</td>
<td>.GI.E..A.</td>
</tr>
<tr>
<td>Rat</td>
<td>.G..E.</td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td>.S...</td>
</tr>
<tr>
<td>Chick</td>
<td>.T..N.</td>
</tr>
<tr>
<td>Alligator</td>
<td>17aa alternative splice</td>
</tr>
<tr>
<td>Xenopus</td>
<td>.T..NE..SA.M.YS.</td>
</tr>
<tr>
<td>Axolotl</td>
<td>.A..NE..A.M.YS.</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>IT.A</td>
</tr>
<tr>
<td>Human</td>
<td>EXON8 ZF2</td>
</tr>
<tr>
<td>Pig</td>
<td>LVRSASETSEKRPFMCAVPQCNKRYFKLISHOMHSRQ7TEGEKPYQCDFKDCCE</td>
</tr>
<tr>
<td>Mouse</td>
<td>I...T..N.</td>
</tr>
<tr>
<td>Rat</td>
<td>I...T..N.</td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td>I...N.</td>
</tr>
<tr>
<td>Chick</td>
<td>Alligator</td>
</tr>
<tr>
<td>Axolotl</td>
<td>Zebrafish</td>
</tr>
<tr>
<td>Human</td>
<td>EXON9 ZF3</td>
</tr>
<tr>
<td>Pig</td>
<td>BRF5RSDQLKPHQRFQHMVDKPQQKTCQKQRKFKSRSRLHKTTHRTHTGTKSEKP</td>
</tr>
<tr>
<td>Mouse</td>
<td>I...T..N.</td>
</tr>
<tr>
<td>Rat</td>
<td>Alligator</td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td>3aa alternative splice</td>
</tr>
<tr>
<td>Chick</td>
<td>Axolotl</td>
</tr>
<tr>
<td>Human</td>
<td>EXON10 ZF4</td>
</tr>
<tr>
<td>Mouse</td>
<td>FSKCRWPSNOQCFARSDELVRRHNNQHQRNMTKQLLAL*</td>
</tr>
<tr>
<td>Rat</td>
<td>.H.</td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td>.Y...K.T.</td>
</tr>
<tr>
<td>Chick</td>
<td>Alligator</td>
</tr>
</tbody>
</table>
for which a comparison was obtained was the zinc fingers for which we already have structural information.

5.2.1.2 Post-translational modification

To investigate whether WTI protein may be post-translationally modified, searches were made for recognition sequences for N-linked glycosylation (Doolittle 1986) and phosphorylation (Kemp and Pearson 1990) sites. No putative glycosylation sites were found in any species, but putative phosphorylation sites were found. Recognition sites found included casein kinase I and II, proline dependent kinase and cAMP dependent kinase. Caution must be observed when drawing conclusions from the presence of recognition sites as it is known that not all sites are recognised and there can be synergy between different sites. Although these sites are present, no phosphorylation of WTI has been observed (Morris et al., 1991). The conservation of these sites was found to be variable with some being present in all species and some in only one.

5.2.1.3 Exon structure

To investigate if the exon structure of WTI is conserved in all vertebrates will require the isolation of genomic sequences. Comparison of the sequences at the known human splice sites in the cDNAs isolated can only give limited information because the most invariant splice site sequences occur in the introns. The first and last two bases of an exon tend to be GT and AG. This is the case in some of the human WTI splice sites, but most are at least purine rich. Table 5.1 shows that most of these sites are conserved. Substitutions which have occurred increase the similarity to the consensus sequence, indicating that the exon structure may well have been conserved, as seen for exons 7-9 in the zebrafish.

5.2.2 5' Non coding region

Multiple transcription start points have been reported in the WTI promoters from mouse and human, approximately 500-400bp upstream of the predicted initiation codon (Pelletier et al., 1991c, Hofmann et al., 1993). Within the 5' non coding region there is still over 80% nucleotide sequence identity between mouse and human (Table 5.2). Alignment with the chick reveals a lower level of conservation but several regions with a higher than background level of identity. These may correspond to protein binding sites (Figure 5.2). WTI has been shown to bind this region in mouse (Rupprecht et al., 1994) but since the true DNA consensus sequence for WTI is unknown comparison cannot reveal whether binding is conserved in chick. The C+G
Table 5.1 Conservation of splice sites in \textit{WT1}

The sequence of the exons at the splice donor and acceptor sites in human \textit{WT1} is compared with the consensus sequence and the sequence in the \textit{WT1} orthologs.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Species</th>
<th>Splice acceptor sequence</th>
<th>Splice donor sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GTA/GAGT</td>
<td>A/CAG</td>
</tr>
<tr>
<td>1</td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GTA/GAGT</td>
<td>GCAA/TCAACG</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chick</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GTA/GAGT</td>
<td>GCTCGCTG</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chick</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GTA/GAGT</td>
<td>CTACACCG</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chick</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alligator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GTA/GAGT</td>
<td>CTCAAGGG</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GTA/GAGT</td>
<td>AGCAGC</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GTA/GAGT</td>
<td>CTCAAG</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chick</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alligator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GTA/GAGT</td>
<td>AGCACACTG</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marsupial mouse</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chick</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alligator</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xenopus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zebrafish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GTA/GAGT</td>
<td>GACAATCACG</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marsupial mouse</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chick</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alligator</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xenopus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zebrafish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GTA/GAGT</td>
<td>CTCAATACAG</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marsupial mouse</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chick</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alligator</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xenopus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zebrafish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>GTA/GAGT</td>
<td>GTAAAAACAA</td>
</tr>
<tr>
<td></td>
<td>Rat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marsupial mouse</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chick</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alligator</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zebrafish</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
content of this region is high (over 70%) in all species; this drops along the sequence falling to 50% in exon 4. The percentage of CpG dinucleotides is normally under-represented in the vertebrate genome due to deamination of methylated cytosine (Bird 1986). In exon 1, CpG dinucleotides have been retained in all species. This would indicate the conservation of the mammalian CpG island in chick and probably transcriptional control of this region.

5.2.3 The predicted initiation codon is conserved

The clone C2.1 (chick cDNA clone - Figure 4.10) extended 370bp 5' to the predicted mammalian methionine initiation codon. Although it has not been demonstrated in vivo that this is the initiation codon the protein observed on Western blots is of the predicted size (Morris et al., 1991, Telerman et al., 1992). Table 5.2 compares the nucleotide identity between species in the coding and non coding regions of exon 1 and shows that the identity between all species increases immediately downstream of the putative initiation codon consistent with this being the initiation codon. The Kozak sequence is the optimum sequence required for initiation of translation in eukaryotes (Kozak 1987). It has been found that this is not the most frequent sequence at translation start sites and A/GNC most commonly precedes the initiation codon (Cavener and Ray 1991). This is not the case in all the WTI orthologs, but this sequence is found two bases further upstream, probably making it a favourable site of translation initiation (Table 5.3). Any upstream methionine initiation codons that are present in human, mouse and chick transcripts are followed by downstream termination codons indicating no N-terminal extended version of WTI can be produced.

5.2.4 Transregulatory domain

5.2.4.1 Homopolymer domains

The transactivating/repressing activity of WTI has been located to the N-terminal 324 amino acids (exons 1-7) using domain swap experiments (Madden et al. 1991). In the amino acid comparison the two most striking inter-species changes in this domain are those affecting two homopolymer sequences (Proline)13 (54-68aa) and (Glycine)5 (19-23aa) (Figure 5.1). The lack of conservation of these regions is especially noticeable in the context of rest of the transregulatory domain where only a few scattered amino acid changes are seen (Figure 5.1).
Figure 5.2
Multiple sequence alignment of the non-coding region of exon 1 using CLUSTAL V
Sequence footprinted by both -/+KTS isoforms are indicated (overlined) as well as regions of greater than average similarity between the three species (highlighted).

<table>
<thead>
<tr>
<th></th>
<th>HUMAN</th>
<th>MOUSE</th>
<th>CHICK</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT1 binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN</td>
<td>AGTG--AAAGCGCCAGGCTT-TGCTTCTCTCTCTCCTGTGGTCAAGGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>AGTG--AAAGCGCCAGGCTT-TGCTTCTCTCTCTCCTGTGGTCAAGGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICK</td>
<td>AGTG--AAAGCGCCAGGCTT-TGCTTCTCTCTCTCCTGTGGTCAAGGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>CHICK</td>
</tr>
<tr>
<td>WT1 binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN</td>
<td>-CGACACACAG-AGGGCGCTGAGGAGGCTGTGGTCAAGGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>-CGACACACAG-AGGGCGCTGAGGAGGCTGTGGTCAAGGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICK</td>
<td>-CGACACACAG-AGGGCGCTGAGGAGGCTGTGGTCAAGGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>CHICK</td>
</tr>
<tr>
<td>WT1 binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN</td>
<td>CTTATTTGAGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>CTTATTTGAGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICK</td>
<td>CTTATTTGAGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>CHICK</td>
</tr>
<tr>
<td>WT1 binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN</td>
<td>GCCGCCAAGACGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>GCCGCCAAGACGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICK</td>
<td>GCCGCCAAGACGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>CHICK</td>
</tr>
<tr>
<td>WT1 binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN</td>
<td>CCAACCAAGACGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>CCAACCAAGACGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICK</td>
<td>CCAACCAAGACGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>CHICK</td>
</tr>
<tr>
<td>WT1 binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN</td>
<td>GCTGCTAGCTGAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>GCTGCTAGCTGAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICK</td>
<td>GCTGCTAGCTGAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>CHICK</td>
</tr>
<tr>
<td>WT1 binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN</td>
<td>GTGCTAGCTGAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>GTGCTAGCTGAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICK</td>
<td>GTGCTAGCTGAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>CHICK</td>
</tr>
<tr>
<td>WT1 binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN</td>
<td>AGTGCTAGCTGAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>AGTGCTAGCTGAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICK</td>
<td>AGTGCTAGCTGAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>CHICK</td>
</tr>
<tr>
<td>WT1 binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN</td>
<td>GTCTGCTAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>GTCTGCTAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICK</td>
<td>GTCTGCTAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>CHICK</td>
</tr>
<tr>
<td>WT1 binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN</td>
<td>AAGTGCTAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>AAGTGCTAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICK</td>
<td>AAGTGCTAGCTGAGGCTTTGGGAAGCTGAGGGCAGGCTTACCCCTCCCTCCTTGAGGA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

147
Table 5.2
Percentage nucleotide identity between WTI orthologs in the non coding and coding regions of exon 1

<table>
<thead>
<tr>
<th></th>
<th>Non coding</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mouse</td>
<td>chick</td>
</tr>
<tr>
<td>human</td>
<td>81.4</td>
<td>45</td>
</tr>
<tr>
<td>mouse</td>
<td></td>
<td>47</td>
</tr>
<tr>
<td>human</td>
<td>87.9</td>
<td>89.7</td>
</tr>
<tr>
<td>mouse</td>
<td>96.1</td>
<td>79.0</td>
</tr>
</tbody>
</table>
| Calculations performed using the gap program from the GCG

Table 5.3
Comparison of the Kozak sequence with the putative WTI initiation codon

<table>
<thead>
<tr>
<th>Kozak</th>
<th>AGCCGCCGCCATGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>CCTCAAGCAGATGG</td>
</tr>
<tr>
<td>Mouse</td>
<td>CCGCACCTCAGTG</td>
</tr>
<tr>
<td>Rat</td>
<td>ACCTCAGTG</td>
</tr>
<tr>
<td>Chick</td>
<td>TGCGAGCAGATGG</td>
</tr>
</tbody>
</table>

The initiation codon is highlighted and the Cavener and Ray (1991) sequence is underlined

Table 5.4 Percentage amino acid composition of the transregulatory domain

<table>
<thead>
<tr>
<th>Species</th>
<th>Exon</th>
<th>% Proline</th>
<th>% Glutamine</th>
<th>% Serine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>1</td>
<td>20.5</td>
<td>9.6*</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>13.8</td>
<td>9.2</td>
<td>13.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0</td>
<td>11.5</td>
<td>3.8</td>
</tr>
<tr>
<td>Chick</td>
<td>1</td>
<td>14.7</td>
<td>10.9*</td>
<td>13.4</td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>13.8</td>
<td>10.8</td>
<td>10.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0</td>
<td>11.5</td>
<td>7.2</td>
</tr>
<tr>
<td>Alligator</td>
<td>2-3</td>
<td>13.8</td>
<td>9.2</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0</td>
<td>11.5</td>
<td>7.2</td>
</tr>
<tr>
<td>Average for</td>
<td>5.6</td>
<td>4.3</td>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td>Vertebrates (Doolittle 1986)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Calculation using only the second half of exon 1
In human exon 1 there is a region of 15 amino acids, 13 of which are proline (the other two being alanine). In mouse and rat this is one proline shorter, but in contrast it is completely absent in the clone C2.1. PCR with primers flanking this region gave a product 54bp smaller from chick genomic DNA and cDNA than from mouse genomic DNA (Figure 5.3a). Direct sequencing of the products confirmed that this was due to the absence of (Proline)_{13} in chick. The presence of this motif was investigated in other mammals besides the primate and rodent orders by PCR. The PCR products in two artiodactyls (pig and cow) were the same size as that in the mouse (Figure 5.3b) and by direct sequencing the proline stretch was found to be present (data not shown).

This stretch of five glycines (rat and human), or six (mouse), is also located in exon 1. In C2.1 the predicted amino acid sequence of this region is completely diverged due to point mutations at non-degenerate sites (Figure 5.1). Direct sequencing of this region from the PCR products above, showed in pig and cow that the glycine region was conserved in these species (data not shown).

A third homopolymer sequence was also present PPPP (Proline)_{4} (114-117aa) in exon 1. In the chick this was not 100% conserved, being APPPP.

The amino acid composition is similar to other transregulatory domains

The proportion of proline in exon 1 is reduced to 15% in chick (mostly due to the deletion of (Proline)_{13}) as compared with 20% in humans, but this is still higher than the average value for vertebrates of 6.1% (Doolittle 1986) (Table 5.4). This is of importance because the high levels of proline and glutamine are thought to be characteristic of some transactivation domains, although the reason for this is not known (Mitchell and Tjian 1989). Although these levels are conserved in WT1 and are more than double the average (Table 5.4) they are lower than the proline/glutamine rich regions of other transcription factors. The transactivation domains of CTF and human and Xenopus Sp1 contain 20-30% proline and 25% glutamine respectively (Mitchell and Tjian 1989). WT1 is more similar to another transcription factor, Krox 20 (EGR2), where the transactivation domain contain 12% proline and 13% serine. Krox 20 also contains a conserved proline stretch, in this case it is seven residues long (Sukhatme et al., 1988, Oxtoby and Jowett 1993). The proline and glutamine rich regions of WT1 overlap but they are not all identical. The proline rich region extends from exon 1 to 3 (amino acids 1-223), but the glutamine rich region covers the C-terminal half of exon 1 to exon 4 (amino acids 73-248). This may point to a modular arrangement of functions.
Figure 5.3
Investigation into the conservation of homopolymer domains in exon 1

a Schematic comparison of the human, mouse and rat WTI amino acid sequence in exon 1 with the sequence derived from the chick clone C2.1. The two homopolymer domains (Glycine)5 and (Proline)13 were found not to be conserved in C2.1.

b Verification that the loss of the homopolymer domains is not a cloning artefact. PCR across the two domains with primers D742 and B621 showed that the band obtained with chick genomic and cDNA templates is the same size as from C2.1. These three bands are all approximately 50bp smaller than that amplified from mouse genomic DNA.

(B621 5' GGCTTGCAGGGTGGGCGC(A/G)TTIGG(A/G)AACA 3')
(D742 5' AGATGGGGTCCGACGTCCGGGACC 3')

c Investigation of exon 1 in other mammals. PCR using pig and cow genomic DNA templates shows that the product is the same size as the product from mouse genomic DNA.
5.2.4.3 Mutations

There are two instances where a somatic mutation in a Wilms' tumour occurs in the transregulatory domain. These are a phenylalanine to serine change at codon 154 and a glycine to aspartic acid change at codon 201 (Park et al., 1993b&c, Table 1.3). Both are non conservative changes with an aliphatic residue being converted to a more hydrophilic residue. These changes may affect protein interactions, disrupt the hydrophobic core, or the folding of the protein. The G to D mutation changes the \textit{in vitro} transrepression activity to transactivation. In both cases these two residues were found to be conserved; suggesting an important role in all species so far examined (Figure 5.1). The effect of the exclusion of exon 2 by alternative splicing in Wilms' tumour cell lines is to covert \textit{WTI} into trancitivator. Unlike wild type \textit{WTI}, exon 2-less transcripts cannot repress the transformed phenotype of a Wilms' tumour cell line (Haber et al., 1993). This therefore implicates exon 2 in both transcription repression and tumorigenesis, consistent with a high level of evolutionary conservation.

5.2.5 The conservation of putative leucine zipper in exon 4

The proposed leucine zipper motif, defined by the presence of leucines every seven residues, and the absence of helix breaking prolines, is encoded by exon 4. It may extend as far as the first proline in exon 6. Leucine zippers are a subset of helices that are amphipathic. The hydrophobicity of one face of the helix favours the interaction with another such helix on the hydrophobic side, forming a structure called a coiled coil. Figure 5.4a shows chick exon 4 and exon 6 as far as the first proline, drawn in a \(\alpha\) helical format (helicalwheel program-GCG). This demonstrates that there is a grouping of hydrophobic residues on one side of the helix, and the amino acid substitutions observed are located outside this. It is also known that the polar residues are important stabilising the contact between the helices (Conway and Parry 1990). This may explain the high level of conservation of the non-hydrophobic residues.

Although many sequences with a heptad repeat of leucines have been proposed to form coiled coil interactions it is doubtful that they all do (Lupas et al., 1991). It is now possible to estimate statistically the likelihood of a coiled coil being formed by comparison with the positions of amino acids with proven coiled coil forming domains using the Coils program (Lupas et al., 1991, and updated 1993, A Lupas, personal communication). The predicted likelihood of a coiled coil forming in this region of the \textit{WTI} protein is very low. This program is known not to detect buried helices or
Figure 5.4 Helical representation of the putative leucine zipper

Sequence of chick and human WT1 covering the putative leucine zipper region, exons 4-6 (221-279aa, human numbering). This was drawn in \( \alpha \) helical format, using the helical wheel programme of GCG. Hydrophobic residues are boxed

a In the -17aa isoform (chick WT1) there is a grouping of hydrophobic residues and a lack of hydrophilic residues on one side of the helix. This is conserved in all species examined. Cross species substitutions are indicated for all orthologs except pig. Most variability is located in the outer ring of the wheel which corresponds to exon 6 sequence (* = exon 6 onwards).

b When the 17aa alternative splice is included in mammals the localisation of hydrophobic residues is changed. (Cross species substitutions shown for the alternative splice only)
those in bundles of more than three in antiparallel orientation. However if these types of helices were present in WTI their availability for the proposed protein-protein interaction function is probably limited.

5.2.6 Exon 5 - The 17aa alternative splice is not conserved in the archosaur lineage

Exon 5 is an alternatively spliced exon in human, mouse (Haber et al., 1991) and rat (Sharma et al., 1992) and the sequence is conserved in those species. In this study its presence has only been observed in the clones obtained from the mammalian species rat and pig (Figure 4.5). This demonstrates that although the PCR method by which the WTI orthologs were isolated did not exclude +17aa isoforms none were obtained from chick or alligator. To investigate whether this alternative splice occurs in the alligator and chick, PCR with primers flanking this region was performed with rat, chick and alligator cDNA template. A band was seen corresponding to the +17aa isoform only in the rat, but in all species the smaller isoform was detected (Figure 5.5). It may be that a low and therefore undetected level of the +17aa isoform is produced. Even this explanation points to a different role to mammals as at least 60% of transcripts in mouse and human are +17aa (Haber et al., 1991).

5.2.7 Divergence and conservation of exon 6

Exon 6 is the first exon within which a clear difference between conserved and unconserved regions is apparent (Figure 5.6). The C-terminal end of exon 6 is perfectly conserved whereas there are many substitutions in the N-terminal end of the exon. The adjacent N-terminal region of exon 7 is also subject to variation, although these are only conservative changes. This may point to an important role for the C-terminal region of exon 6, although it contains no known motifs.

The first point mutation outside the zinc fingers in WTI was observed in a benign mesothelioma (Park et al., 1993a). The tumour-specific homozygous A to G (1195nt) transition resulted in a serine to glycine (273aa) substitution which is considered to be a conservative change (Dayhoff et al., 1972). This particular mutation was shown to have the capacity, in vitro, to change the transrepression activity of WTI to transactivation. This is surprising considering that in all species except the eutherian mammals, this position is found to be an asparagine. An asparagine to serine change is considered a non conservative substitution. This
Figure 5.5

Investigation into the conservation of the 17 amino acid alternative splice-exon 5

a Schematic representation of the + and - 17aa isoforms of WTI found in mammals showing the position of the PCR primers used.

b Amplification across exons 4-9 in chick, alligator and rat cDNA. Two bands 51bp apart corresponding to +/- 17aa isoforms are observed in rat. But in chick and alligator no larger band is detected.
(D797 5' AGCTTGAATGCATGACATGGAA 3')
(D798 5' TGGGTGTGTATTCTGTATAGGG 3')
substitution may have become fixed because other substitutions in this region compensate for, or neutralise, this change, or there is a functional difference between the species.

The region has also been implicated in type II RNA editing in rat and human (Sharma et al., 1994). A T to C (1220nt) transition was found in the some mRNA transcripts, resulting in a leucine to proline (281aa) substitution, that was not encoded by the genomic DNA. Leucine was found to be conserved at this position in all species so far examined, including the rat. To try to assess the likelihood that editing at this position is conserved in other species the conservation of the nucleotide sequence was examined (Figure 5.6). RNA editing in mammals has been most closely studied in apolipoprotein B (Chan 1993). Dissection of the sequences required for editing has shown that an 11bp sequence, 5bp downstream from the edited position, is required and is not species-specific (Backus and Smith 1991). A species-specific sequence has been inferred further downstream. Downstream of the edited site in WTI there are certain conserved similarities to the apolipoprotein B non species-specific editing box. In alligator and chick immediately downstream from the affected nucleotide there is a 6 nucleotide insertion. This insertion results in a tyrosine, serine insertion in the gene product (Figure 5.6). Therefore the editing box in chick and alligator is in a different location relative to the putative edited nucleotide and so it is unlikely that the same nucleotide is edited in the reptilian lineage. The effect of a leucine to proline substitution in alligator and chick may be different to mammals because of presence of the YS insertion in the gene product. In the chick the apolipoprotein form produced by editing has not been detected; so editing of apolipoprotein B is thought not to be conserved in chick (Tarugi et al., 1990).

5.2.8 The zinc finger region is the most highly conserved region of WTI

The region encoding the first three zinc fingers was the region for which the greatest number of sequences was obtained and therefore has the greatest depth to the comparison. In exons 7-10, which encode the zinc fingers of WTI, only 14 changes were observed in the 130 amino acid residues compared. Of these, seven were observed in the part of exon 7 which precedes ZF1, and these were conservative changes (Dayhoff et al., 1972) or a deletion/insertion (Figure 5.1). Five of the seven substitutions found within the fingers were located either in the predicted β loop between the two zinc chelating cysteines or in the α helix between the two zinc chelating histidines (Figure 5.7). Another residue substituted in ZF2 of the zebrafish
Figure 5.6 Mutation, editing and conservation of exon 6

The changes in the exon 6 sequence due to mutation and editing in mammals are compared with substitutions in other species. The serine residue mutated in a benign cystic mesothelioma, causing a loss of transrepressor activity and gain of transactivator activity is not conserved outside of the eutherian mammals. The edited nucleotide in human and rat causes a leucine to proline change at a position conserved in all species. The nucleotide sequence does show some similarity to the 3' non species specific editing box in apolipoprotein but in alligator and chick this site is no longer 5bp downstream due to the insertion of 6bp directly 3' to the edited nucleotide.
WT1 is immediately after the second invariant cysteine in the β sheet. In the X-ray co-crystal structure of EGR1(Zif 268) all these residues are away from the DNA binding end of the finger (Figure 1.2b) and therefore probably have a limited role in binding to DNA (Pabo and Pavletich 1991). EGR1 mutations at these sites had no effect on DNA binding (Wilson et al., 1992).

The one substitution found in Xenopus was in the linker region between ZF2 and 3. It is conservative substitution of one hydrophobic amino acid for another (valine for isoleucine). In a study of mutations which would change the function of the zinc finger proteins EGR1 and TFIIIA, changes at this position were found to have no effect (Wilson et al., 1992, Choo and Klug 1993). This is in contrast to the neighbouring lysine residue where mutation inhibits DNA binding activity (Wilson et al., 1992). All the substitutions found are therefore located in sites already thought to have lesser importance in the function of the zinc fingers. This is in contrast to the positions at which pathogenic mutations have been found, which are all conserved residues.

One other change in the fingers apart from those mutations found in DDS and Wilms' tumours was found in one of the first WT1 transcripts to be isolated, WT33 (Call et al., 1990). This is a serine to phenylalanine change (365aa) in ZF2 in the C-terminal end of the β sheet (Figure 5.7); this is the part of the finger involved in DNA binding and the substitution has been shown to affect binding to some DNA target sequences (Bickmore et al., 1992). It is not known whether this is a pathogenic mutation or a polymorphism, but as it has only been observed once it is unlikely to be a polymorphism. The library from which it was isolated was from a pre-B cell leukaemia cell line and so this change may be a pathogenic mutation involved in tumour progression (Bickmore et al., 1992). This is also suggested because this position has been shown to be a conserved in WT1 and in other homologous finger proteins.

The contribution made by ZF1 to the binding of WT1 to its target DNA sequence is not known but a constitutional mutation of one of the zinc chelating cysteines results in DDS (Bruening et al., 1992). The complete conservation through 400Myr, since the divergence of the fish lineage, supports an important function for ZF1. Although no missense mutations in ZF4 have been reported, this finger is conserved to the same degree as the other three fingers, indicating a functional role.
Figure 5.7 Schematic representation of the four zinc fingers of WTI.

The region cloned and sequenced in each species is shown and any substitutions relative to the human sequence are indicated. The positions of the missense DDS mutations are also indicated as are the important residues in DNA binding; inferred from EGR1 and Sp1 fingers (Pabo and Pavletich 1991, Kriwacki et al., 1992). The substitutions found in the different species can be seen not to coincide with any mutated or any other functionally important sites.
Amino acid substitutions

- Zebrafish
- Axolotl
- Xenopus
- Rat
- Mouse

- position of DDS misense mutations in humans
- no. of spots indicates number of different changes

- phosphate binding
- base binding
- helix region

Zinc Finger 1
- position of DDS misense mutations in humans
- no. of spots indicates number of different changes

Region cloned
- Mouse
- Rat
- Zebrafish
- Xenopus
- Axolotl
5.2.9 The KTS alternative splice has been conserved through vertebrate evolution

The role of the KTS alternative splice is not well understood. Previously the presence of +KTS isoform had been reported in mammalian species (human, mouse and rat). Transcripts containing the KTS isoform have been isolated from the non mammalian species, marsupial mouse, chick and alligator, and in the latter two both + and -KTS transcripts have been observed (Figure 4.5, Figure 4.11). In the zebrafish genomic clone, Z1a, the 9bp which encode the KTS are present at the 3' end of ZF3 and, in contrast to the lack of conservation in rest of the introns, both alternative splice donor sites are perfectly conserved between human and zebrafish. The conservation extends well past the minimal consensus sequence, making it very likely that +/-KTS isoforms are produced (Figure 4.12).

5.2.10 The C-Terminal region is not changed in mouse

C-terminal to the second conserved histidine of ZF4 there are only 11 amino acids before the stop codon in mammals, marsupials and alligator. In the published mouse and rat sequences (Buckler et al., 1991, Sharma et al., 1992) there is a two amino acid difference to the human sequence, QL becoming HV (448-9aa). This is not seen in either alligator or marsupial sequences and is attributable to an inversion from GC to CG (1715-1716nt) in human to mouse. To investigate if this a real sequence difference a mouse cDNA clone KS/2 (gift Buckler et al.) was sequenced and found to resemble the human (Figure 5.8). This has also been independently confirmed by Melissa Little (personal communication). This would explain why monoclonal antibodies raised against this region of the human sequence also bind the mouse WT1 protein (V. van Heyningen personal communication). It is highly likely that the original sequencing error has been perpetuated in the sequence analysis of the rat. All calculations included in this study therefore have been based on the C-terminal end of mouse and rat WT1 resembling human, marsupial and alligator.

5.2.11 3' Non coding region

Clones which extended past the stop codon were obtained for marsupial mouse and alligator. In the 3' untranslated region the level of identity between the clones became much lower, and an increase in the AT richness and repetitive character of the sequence was seen (Table 5.5, Appendix A). In the case of the marsupial mouse the
Figure 5.8 Sequence of the mouse Wt1 cDNA clone, KS/2.

The sequencing gel of the mouse clone KS/2 shown above differs from the published mouse sequence (Buckler et al., 1991) by the inversion of a GC to a CG and therefore a two amino acid difference QL to HV. The mouse sequence therefore does not differ from the human in the C-terminal region as previously supposed.
Table 5.5
Percentage similarities between coding and non coding regions of exon 10

Coding

<table>
<thead>
<tr>
<th></th>
<th>mouse</th>
<th>rat</th>
<th>marsupial mouse</th>
<th>alligator</th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>95.0</td>
<td>92.6</td>
<td>88.5</td>
<td>83.6</td>
</tr>
<tr>
<td>mouse</td>
<td>95.9</td>
<td></td>
<td>86.9</td>
<td>81.3</td>
</tr>
<tr>
<td>rat</td>
<td></td>
<td>84.4</td>
<td>80.3</td>
<td></td>
</tr>
<tr>
<td>marsupial mouse</td>
<td></td>
<td></td>
<td>84.4</td>
<td></td>
</tr>
</tbody>
</table>

Non coding

<table>
<thead>
<tr>
<th></th>
<th>mouse</th>
<th>rat</th>
<th>marsupial mouse</th>
<th>alligator</th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>72.0</td>
<td>73.4*</td>
<td>45.5*</td>
<td>45.2*</td>
</tr>
<tr>
<td>mouse</td>
<td>91.4*</td>
<td></td>
<td>48.2*</td>
<td>43.6*</td>
</tr>
<tr>
<td>rat</td>
<td>38.7*</td>
<td></td>
<td>41.3*</td>
<td></td>
</tr>
<tr>
<td>marsupial mouse</td>
<td></td>
<td></td>
<td>39.8*</td>
<td></td>
</tr>
</tbody>
</table>

Calculations performed using the Gap program from GCG.
Figures calculated using of the corrected mouse and rat sequences (see text for details)
* indicates incomplete sequences compared
entire 3' non coding was isolated as polyA tail was observed at the 3' end of the clone. The marsupial mouse untranslated region was approximately 1.3kb long (only 400bp were sequenced) which is the same as the length of this sequence in mouse and human. In the 3' untranslated region of the marsupial mouse the similarity to the human, mouse and rat regions was found to be as low as the alligator (Table 4.5). At the 3' end of the marsupial sequence a higher degree of similarity to the eutherian mammals was seen including a AUUAAA sequence approximately 15bp away from the beginning of the polyA tail (in bold in Appendix A). This is a weaker polyadenylation signal than the canonical AAUAAA (Proudfoot 1994) but its position and conservation in marsupial, human and mouse suggests it is the polyadenylation signal recognised. In humans and mouse a polyadenylation signal (AAUAAA) is located approximately 900bp from the 3' end of the mRNA. In alligator this sequence was repeated twice (890-899bp in AL1), only 20bp downstream from the stop codon, suggesting that in certain circumstances shorter transcripts could be produced if these sequences were recognised.

5.3 Evolutionary analysis of WTI

The main objectives of the comparison of WTI orthologs was to gain information at the structural and functional level, rather than to address the issues of rate of evolution and vertebrate phylogeny. Analysis of changes in the nucleotide sequence of WTI with time was performed mainly to investigate the reason for the high level of conservation of the protein sequence. At the same time it was tested if the data from these sequences were consistent with present notions of evolution and vertebrate phylogeny.

5.3.1 Rate of sequence evolution of WTI

5.3.1.1 Introduction

The structural analysis of WTI has shown that the predicted amino acid sequence has diverged very little in particular in the zinc finger domain. At the nucleotide level comparison of the of the coding versus the non coding regions (introns, 5' and 3' non coding regions, (Tables 5.2&5.5)) shows that the rate of change of the non coding region must be very much higher than the coding region. It is now known that the rate of change of coding sequences is dependent on both the actual rate of mutation and the amount of purifying selection exerted on the gene. For
protein coding sequences most of the purifying selection will operate at the level of
the protein produced. Mutations which affect the nucleotide sequence without
changing the encoded protein (synonymous sites) are effectively 'silent' and can be
fixed in a population by such means as genetic drift. However these substitutions can
have an effect at the level of GC content, codon preference, neighbouring base
mutational effects, and RNA secondary structure (Eyre-Walker 1991). Therefore to
quantify the evolution of a nucleotide sequence it is better to compare the 'silent' and
non 'silent' sites separately rather than comparing the sequences simply on a
percentage basis. A percentage comparison does not compensate for multiple,
backwards and convergent substitutions. These can be corrected for, to some extent,
by equations such as the Jukes and Cantor formula (Jukes and Cantor 1969).

Several programs for comparing the rate of mutation at synonymous and
nonsynonymous sites have been produced. The method published by Li et al., 1985
and updated in 1993 (Li 1993) was used to calculate the rate of synonymous
substitutions per synonymous site (K_S) and the rate of nonsynonymous substitutions
per nonsynonymous site (K_A) for WTI. The rate of nonsynonymous substitutions can
be compared to the rate of synonymous substitutions by the formula K_A/K_S. If there
is selection acting on the protein sequence the value for K_A will be much lower than
the value for K_S.

The program classifies each site in the two sequences compared into 1) non-
degenerate, 2) two-fold degenerate or 3) four-fold degenerate, depending on whether
nucleotide substitutions at a site will change the amino acid sequence 1) always (non-
degenerate), 2) one out of three times (two-fold) or 3) never (four-fold degenerate).
The values obtained are then averaged between the two sequences. The sequences
are then compared and the changes classified according to the above scheme. At two-
fold sites transitions are taken as degenerate changes and transversions as non-
degenerate. In the case of multiple substitutions in a codon the pathway within which
the least number of non-degenerate changes would have occurred is given the higher
weighting.

5.3.1.2 Analysis of WTI

The sequence information obtained for the WTI orthologs did not cover
identical regions. It was suspected that analysing different regions for different
species may well produce a bias in the figures, in particular for those clones where
only the highly conserved zinc fingers had been obtained. Therefore exons 7-9 were
Table 5.6 Figures obtained for $K_S$ and $K_A$ using Li's program

### a Exons 7-9

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>M</th>
<th>R</th>
<th>P</th>
<th>MM</th>
<th>C</th>
<th>AL</th>
<th>X</th>
<th>AX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>22</td>
<td>24</td>
<td>10</td>
<td>55</td>
<td>58</td>
<td>60</td>
<td>114</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6)</td>
<td>(6)</td>
<td>(3)</td>
<td>(13)</td>
<td>(13)</td>
<td>(12)</td>
<td>(34)</td>
<td>(18)</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>0.5</td>
<td>0.0</td>
<td>5</td>
<td>22</td>
<td>58</td>
<td>88</td>
<td>60</td>
<td>105</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
<td>(0.0)</td>
<td>(2)</td>
<td>(6)</td>
<td>(12)</td>
<td>(19)</td>
<td>(12)</td>
<td>(30)</td>
<td>(23)</td>
</tr>
<tr>
<td>Rat</td>
<td>0.5</td>
<td>0.0</td>
<td>26</td>
<td>62</td>
<td>87</td>
<td>67</td>
<td>97</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
<td>(0.0)</td>
<td>(7)</td>
<td>(13)</td>
<td>(19)</td>
<td>(14)</td>
<td>(25)</td>
<td>(24)</td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>0.0</td>
<td>0.5</td>
<td>0.5</td>
<td>62</td>
<td>84</td>
<td>80</td>
<td>111</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(14)</td>
<td>(19)</td>
<td>(19)</td>
<td>(32)</td>
<td>(44)</td>
<td></td>
</tr>
<tr>
<td>Marsupial</td>
<td>1.3</td>
<td>1.7</td>
<td>1.6</td>
<td>1.3</td>
<td>85</td>
<td>79</td>
<td>112</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>(0.8)</td>
<td>(0.9)</td>
<td>(0.9)</td>
<td>(0.7)</td>
<td>(19)</td>
<td>(16)</td>
<td>(34)</td>
<td>(19)</td>
<td></td>
</tr>
<tr>
<td>Chick</td>
<td>1.7</td>
<td>2.1</td>
<td>2.0</td>
<td>1.7</td>
<td>0.4</td>
<td>25</td>
<td>79</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.9)</td>
<td>(0.9)</td>
<td>(0.9)</td>
<td>(0.8)</td>
<td>(0.4)</td>
<td>(6)</td>
<td>(20)</td>
<td>(14)</td>
<td></td>
</tr>
<tr>
<td>Alligator</td>
<td>1.7</td>
<td>2.1</td>
<td>2.0</td>
<td>1.6</td>
<td>0.4</td>
<td>0.0</td>
<td>99</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.9)</td>
<td>(0.9)</td>
<td>(0.9)</td>
<td>(0.8)</td>
<td>(0.4)</td>
<td>(0.0)</td>
<td>(25)</td>
<td>(13)</td>
<td></td>
</tr>
<tr>
<td>Xenopus</td>
<td>0.9</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.7)</td>
<td>(0.7)</td>
<td>(0.7)</td>
<td>(0.7)</td>
<td>(0.8)</td>
<td>(0.7)</td>
<td>(0.7)</td>
<td>(19)</td>
<td></td>
</tr>
<tr>
<td>Axolotl</td>
<td>0.9</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.9)</td>
<td>(0.8)</td>
<td>(0.8)</td>
<td>(0.8)</td>
<td>(0.8)</td>
<td>(0.8)</td>
<td>(0.8)</td>
<td>(1.1)</td>
<td></td>
</tr>
</tbody>
</table>

### b Exons 3-10

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>M</th>
<th>R</th>
<th>P</th>
<th>MM</th>
<th>C</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>35.3</td>
<td>34.2</td>
<td>72.6</td>
<td>58</td>
<td>114</td>
<td>60</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>(5.7)</td>
<td>(5.4)</td>
<td>(11.3)</td>
<td>(13)</td>
<td>(34)</td>
<td>(12)</td>
<td>(18)</td>
</tr>
<tr>
<td>Mouse</td>
<td>1.8</td>
<td>10.7</td>
<td>81.6</td>
<td>88</td>
<td>60</td>
<td>105</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
<td>(2.7)</td>
<td>(11.7)</td>
<td>(19)</td>
<td>(12)</td>
<td>(30)</td>
<td>(23)</td>
</tr>
<tr>
<td>Rat</td>
<td>1.7</td>
<td>0.8</td>
<td>86.2</td>
<td>94.7</td>
<td>95.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
<td>(0.4)</td>
<td>(12.3)</td>
<td>(13.6)</td>
<td>(13.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>1.6</td>
<td>3.0</td>
<td>2.4</td>
<td>84.0</td>
<td>89.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
<td>(0.8)</td>
<td>(0.8)</td>
<td>(12.9)</td>
<td>(14.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>3.3</td>
<td>3.6</td>
<td>3.9</td>
<td>1.2</td>
<td>84.0</td>
<td>32.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.8)</td>
<td>(0.9)</td>
<td>(0.9)</td>
<td>(0.6)</td>
<td>(12.9)</td>
<td>(5.3)</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2.5</td>
<td>2.7</td>
<td>3.1</td>
<td>1.3</td>
<td>1.0</td>
<td>32.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.7)</td>
<td>(0.7)</td>
<td>(0.8)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(5.3)</td>
<td></td>
</tr>
</tbody>
</table>

Estimation of the rate of change of WT1
Calculations were performed for two regions of WT1 using the program (Li 1993) for estimating the number of synonymous changes per synonymous site ($K_S$) and the number of nonsynonymous changes per nonsynonymous site ($K_A$). (Standard errors are in parenthesis)
Table 5.7 Comparison of $K_S$ and $K_A$ values for WTI with other genes

a) Comparison of $K_S$ and $K_A$ between mouse and rat

<table>
<thead>
<tr>
<th>Gene</th>
<th>$K_S$</th>
<th>$K_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chaperonin</td>
<td>0.077</td>
<td>0.0007</td>
</tr>
<tr>
<td>αβ Crystallin</td>
<td>0.182</td>
<td>0.0000</td>
</tr>
<tr>
<td>Nucleolin</td>
<td>0.110</td>
<td>0.0281</td>
</tr>
<tr>
<td>GADPDH</td>
<td>0.167</td>
<td>0.0104</td>
</tr>
<tr>
<td>Average</td>
<td>0.144</td>
<td>0.0148</td>
</tr>
<tr>
<td>WTI (exons 7-9)</td>
<td>0.050</td>
<td>0.0000</td>
</tr>
<tr>
<td>WTI (exons 3-10)</td>
<td>0.107</td>
<td>0.0080</td>
</tr>
</tbody>
</table>

Data taken from Li 1993

b) Comparison of $K_A/K_S$ in the mammalian lineage

<table>
<thead>
<tr>
<th>Gene</th>
<th>$K_A/K_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aldolase A</td>
<td>0.019</td>
</tr>
<tr>
<td>Thyrotropin</td>
<td>0.017</td>
</tr>
<tr>
<td>ZFY</td>
<td>0.190</td>
</tr>
<tr>
<td>Histone 4</td>
<td>0.000</td>
</tr>
<tr>
<td>β globin</td>
<td>0.262</td>
</tr>
<tr>
<td>IFN γ</td>
<td>0.324</td>
</tr>
<tr>
<td>LDHA</td>
<td>0.040</td>
</tr>
<tr>
<td>Ig V_H</td>
<td>0.189</td>
</tr>
<tr>
<td>Insulin</td>
<td>0.323</td>
</tr>
<tr>
<td>IGF 2</td>
<td>0.225</td>
</tr>
<tr>
<td>WTI (exons 7-9)</td>
<td>0.045</td>
</tr>
<tr>
<td>WTI (exons 3-10)</td>
<td>0.051</td>
</tr>
</tbody>
</table>

Data taken from Li et al., 1985* and Li 1993

*Figures calculated using old method which underestimates $K_S$ and overestimates $K_A$ to some extent.
compared for all the orthologs isolated (Table 5.6a) and exons 3-10 compared in the mammalian and reptilian lineages (Table 5.6b).

In general the increase in evolutionary distance correlated well with an increase in the values of $K_A$ and $K_S$. The only major inconsistency was that the value for $K_S$ in *Xenopus* was much higher than for axolotl, although both are amphibians and therefore similar values would have been expected. This could be owing to the small amount of sequence available and the large standard errors. The $K_A$ values for *Xenopus* and axolotl were lower than for the reptiles although it is believed that amphibians diverged from common ancestors 60Myr earlier (Benton 1990). This is probably because the N-terminal sequence of exon 7 could not be included for *Xenopus* and axolotl.

For exons 7-9 $K_A$ approaches zero as would be expected. The standard error therefore was very high (Table 5.6a). For exons 3-10 $K_A$ was higher than for exons 7-9, consistent with a larger number of amino acid substitutions. In both regions the value for $K_S$ was found to be much higher than the value for $K_A$, thus $K_A/K_S$ is very low, which is consistent with purifying selection acting at the level of the protein sequence. The data obtained for *WT1* were compared with other proteins, the $K_A$, $K_S$ and $K_A/K_S$ values were in the same range as those calculated for other proteins (Table 5.7).

To investigate the relationship of $K_S$ with time, the values for $K_S$ within each clade were averaged and these values plotted against the estimated divergence times of the groups (Benton 1990). The relationship was found to show a linear trend (Figure 5.9). The rate of synonymous substitutions per site per year was found to be similar for both regions examined at $1-1.25 \times 10^{-9}$ using the formula $rate = K_S/2T$ ($T$=divergence time) (Li and Graur 1991). This is lower than values obtained previously ($1.7-11.8 \times 10^{-9}$) (Li *et al.*, 1985) but the updated method (Li 1993) is known to give a lower, more realistic, value for $K_S$. It was noticed from these graphs that the values obtained for comparisons with the chick and alligator sequences lay below the line (Figure 5.9). The $K_S$ and $K_A$ values were subjected to a relative rate test (Li and Graur 1991) and this suggested that the rate of evolution of *WT1* may be slower in the reptilian lineage. This could possibly be an artefact owing to a small data set. To discover the real reason behind the differences in the apparent rates a more extensive sequence data set will be required.
Figure 5.9 Graphs plotting $K_S$ against divergence time

The $K_S$ values (number of synonymous changes per synonymous site) for each species was calculated (Table 5.6) and these were averaged within each group; mammals (eutheria), marsupials, archosaur lineage (birds and some reptiles) and amphibians. The $K_S$ values were then plotted against the divergence times for amphibians from tetrapods, archosaurs from the mammalian lineage, aves from crocodilians, eutheria from metatheria, mammalian radiation and mouse from rat.

No definitive conclusions can be drawn from this data as the errors are high, but points including data from the chick and alligator tended to lie below the line, maybe indicating a non constant rate of evolution of the nucleotide sequence of $WTI$.

a Exons 7-9

b Exons 3-10
5.3.2 Phylogenetic analysis

In order to calculate the divergence between sequences, from which to derive a phylogenetic relationship, multiple alignment of the nucleotide sequence of exons 7-9 was performed using the Clustal V program. The sequences are initially aligned using a fast comparison of the sequences, by producing k-tuple scores for the sequences, plotting these on a dot matrix and finding the diagonals with the most matches. Using the similarity scores a dendrogram was produced using the UPGMA (unweighted pair group method with arithmetic mean) method. The sequences were then aligned starting with the most closely related sequences. This alignment can be used to produce an unrooted phylogenetic tree using the neighbour joining method of Saitou and Nei 1987 (Figure 5.10). This is a distance method, as opposed to a parsimony method, where the percentage divergence is calculated from the alignment and then used to produce the tree. The tree was bootstrapped 500 times (i.e. samples of sequence were taken and from them a tree calculated, the number of times the same internal branches are found gives the degree of confidence in the tree). The weakest parts of the tree were found to correspond to the non mammalian branches. The major inconsistency with palaeontological relationships (Figure 1.7) was also found in this region. The reptiles are shown to form a clade with amphibians rather than with mammals (the amniotes). In addition, the axolotl sequence was found to diverge from the reptiles after *Xenopus*, whereas they would be expected to group together. The figures for the percentage divergence from which the tree was drawn, were plotted against time. A linear increase in divergence with time was seen (Figure 5.11). It was found that the values for the reptilian lineage were again relatively lower than expected, perhaps explaining why the amphibians and reptiles grouped together in the tree.
Figure 5.10 Unrooted phylogenetic tree for exons 7-9 of *WTI*.
This was produced by the neighbour joining method using a CLUSTALV alignment. Branch lengths are drawn approximately to scale and represent 10x percentage divergence. The level of confidence in each branch point as derived from bootstrapping is also indicated.

% occurrence in bootstrap samples

<table>
<thead>
<tr>
<th></th>
<th>97</th>
<th>99</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pig</td>
<td>1.4</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>2.4</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Marsupial mouse</td>
<td>5.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>3.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Axolotl</td>
<td>5.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chick</td>
<td>3.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alligator</td>
<td>3.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zebrafish</td>
<td>13.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xpropus</td>
<td>8.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chick</td>
<td>3.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alligator</td>
<td>13.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zebrafish</td>
<td>3.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5.11 Graph of divergence of WTI against divergence time

Graph plotting percentage divergence, calculated using CLUSTAL V, against divergence time. An approximately linear increase in divergence with time was found. Outlying points were again found to be for reptilian species.
Chapter 6

Investigation of the Expression of WTI
6.1 Introduction

The cloning of WTI orthologs in the different classes of vertebrate will allow the investigation of the similarities and differences of WTI gene expression among them. Knowledge of normal expression patterns is an initial requirement before manipulation of a system can be attempted. Northern blotting and in situ hybridisation techniques were therefore used to investigate the size and expression pattern of WTI mRNA in several of the species from which WTI had been isolated so these could then be compared with that of mammals. The most extensive study was performed in chick and alligator embryos. The development of the genitourinary system in the reptilian lineage is of interest because of its intermediate nature between that of the mammals and lower vertebrates. In alligators the metanephros does form but contains relatively few nephrons while a functional mesonephros persists into post hatching stages (Forbes 1940). In reptiles, but not birds, the mechanism of sex determination is of interest because it is temperature dependent as opposed to being genetically determined (Ferguson and Joanen 1982). This mechanism is under investigation in alligators as it is a solely temperature dependent mechanism. In other reptiles there is also a genetic element. The cloning and determination of the function of genes involved in the early developing gonad will help elucidate this.

6.2 Northern analysis of WTI mRNA

Total RNA was prepared from the genitourinary systems of pig, chick, alligator, and Xenopus. In order to determine the size of WTI mRNA transcripts in these species this RNA was probed with WTI. RNA was resolved on a denaturing agarose gel, and then transferred by vacublotting onto Hybond N (Amersham). The filters were then hybridised at high stringency with the appropriate homologous probe (Figures 6.1&6.2). Details of RNA samples and probes used are contained in Table 4.1 and Figure 4.5.

The major transcript detected in chick, alligator, and pig was approximately 3kb, the same size as that observed in human, mouse, and rat (Call et al., 1990, Buckler et al., 1991, Sharma et al., 1992). An additional 2kb transcript was detected in chick RNA obtained from the complete developing genitourinary system of 11 day embryos, but not from RNA isolated from the mesonephros alone (Figure 6.1).

Total RNA from both whole Xenopus laevis tadpoles (stages 45-46) and mesonephros (stage 55) was probed with XWTI. Consistent with a higher level of
Figure 6.1 Detection of WTI mRNA in pig, chick and alligator

<table>
<thead>
<tr>
<th>Lane</th>
<th>Species</th>
<th>Source of RNA</th>
<th>Organ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pig</td>
<td></td>
<td>Juvenile Testis</td>
</tr>
<tr>
<td>2</td>
<td>Chick</td>
<td>11d genitourinary system</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Chick</td>
<td>13d mesonephros</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Alligator</td>
<td>St. 20 male gonad (33°C)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Alligator</td>
<td>St. 20 female gonad (30°C)</td>
<td></td>
</tr>
</tbody>
</table>

Autoradiograph of a northern blot after high stringency hybridisation with the homologous PCR clones to total RNA from pig, chick and alligator. 3kb bands are present in all tracks demonstrating the conservation of the length of the mRNA transcript in mammals, birds and reptiles. A 2kb band is also detected in the sample from the 11-day chick genitourinary system.
Figure 6.2 Detection of *Xenopus WTI* mRNA

<table>
<thead>
<tr>
<th>Lane</th>
<th>Source of RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>St. 55 mesonephros</td>
</tr>
<tr>
<td>2</td>
<td>St. 45 tadpole</td>
</tr>
<tr>
<td>3</td>
<td>St. 45 tadpole</td>
</tr>
<tr>
<td>4</td>
<td>St. 46 tadpole</td>
</tr>
<tr>
<td>5</td>
<td>St. 46 tadpole</td>
</tr>
</tbody>
</table>

PhosphorImager analysis of a northern blot of total RNA from whole tadpole (St. 45-46) and mesonephros (St. 55) after high stringency hybridisation to *XWTI* (PCR clone). *XWTI* is only detected in the sample from mesonephros and the transcript size is 5kb.
WT1 expression in the mesonephros than elsewhere a 5kb transcript was detected only in mesonephric RNA (Figure 6.2).

6.3 In situ hybridisation

Recently, the technique of whole-mount in situ hybridisation has been developed which allows the expression pattern of a gene to be analysed in the whole embryo as opposed to having to reconstruct this from sections. The embryos can then be sectioned to identify more closely the sites of the expression. In situ hybridisations on sections can be used to obtain better histological detail than in whole-mounts. Whole-mounts also enable unexpected sites of expression to be identified which may not have been included in sections.

Antisense digoxigenin-dUTP labelled riboprobes were prepared from chick and alligator PCR clones CWTL and AWTI (probes described in Figure 4.5) by in vitro transcription, using the Boehringer Mannheim kit. Chick and alligator embryos were collected at various timepoints after incubation at the appropriate temperature. Albino Xenopus laevis embryos between stages 14-39 were generously donated by Richard Harland. Embryos were dissected away from the surrounding membranes, fixed, and hybridised with the appropriate riboprobe (as described in Materials and Methods). The signal was visualised by binding an alkaline phosphatase conjugated antidigoxigenin antibody to the annealed probe. When NBT (nitroblue tetrazolium salt) and BCIP (5-bromo-4-chloro-3-indolyl phosphate) were added, they were catabolised by the alkaline phosphatase to give a insoluble blue precipitate at sites of mRNA expression. To detect any non-specific signal produced due to trapping of the probe (in particular in the intraembryonic spaces), a sense riboprobe or vector-only riboprobe was hybridised in parallel to the antisense riboprobe.

6.3.1 Comparison of the developmental stages of amniotes

A guide to the equivalent stages of human, mouse, chick and alligator development based on somite number and limb development and other developmental features such as branchial arches, is provided in Table 6. Exact whole embryo comparison between species let alone across classes is impossible as different structures are found to develop at different stages and at different rates. This changing of developmental timing and rates (heterochrony) is probably one of the
Table 6 Development of the chick showing comparisons with human, mouse and alligator.

Comparisons based on somite number, limb development or branchial arch or other recorded features

<table>
<thead>
<tr>
<th>Chick stage/age</th>
<th>Chick somite number</th>
<th>Alligator equivalent general features</th>
<th>Mouse equivalent somites/limbs</th>
<th>Human equivalent</th>
<th>Chick developmental events</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Prestreak (no primitive streak)</td>
</tr>
<tr>
<td>St. 2/6-7h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Initial streak</td>
</tr>
<tr>
<td>St. 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Intermediate streak (primitive streak extending)</td>
</tr>
<tr>
<td>St. 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Definitive streak (primitive at maximum length)</td>
</tr>
<tr>
<td>St. 5</td>
<td></td>
<td>St. 11/7.5d</td>
<td></td>
<td></td>
<td>Head process appears</td>
</tr>
<tr>
<td>St. 6/23-24h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Head fold forms</td>
</tr>
<tr>
<td>St. 7</td>
<td>1 somite</td>
<td>St. 12/8d</td>
<td></td>
<td></td>
<td>Nephrogenic cord forms, mesonephros potentiality determined</td>
</tr>
<tr>
<td>St. 8</td>
<td>4 somites</td>
<td>St. 13/8.5d</td>
<td></td>
<td></td>
<td>Neurulation</td>
</tr>
<tr>
<td>St. 9</td>
<td>8 somites</td>
<td>Eggs laid</td>
<td>St. 13/8.5d</td>
<td></td>
<td>Pimordia of the pronephric tubules first appear as paired thickenings in the intermediate mesoderm in the region of somites 5-8</td>
</tr>
<tr>
<td>St. 10</td>
<td>10 somites</td>
<td>St. 14/9d</td>
<td></td>
<td></td>
<td>Pronephric buds grow in a dorsolateral direction</td>
</tr>
<tr>
<td>St. 11</td>
<td>13 somites</td>
<td>St. 1/1</td>
<td>St. 14/9d</td>
<td></td>
<td>Pronephric duct has lumen</td>
</tr>
<tr>
<td>St. 12-13</td>
<td>16-19 somites</td>
<td></td>
<td></td>
<td></td>
<td>Closure of anterior neuropore</td>
</tr>
<tr>
<td>St. 14/50-53h</td>
<td>21 somites</td>
<td>St. 2/2</td>
<td>St. 15/9.5d</td>
<td>24d</td>
<td>Degeneration of pronephros as far as somite 11</td>
</tr>
<tr>
<td>St. 15</td>
<td>24 somites</td>
<td>St. 3/3</td>
<td></td>
<td>25-26d</td>
<td>11 pairs of pronephric tubules 5-16 somite region limb primordia-prospective limb areas flat not yet determined</td>
</tr>
<tr>
<td>St. 16</td>
<td>26-28 somites</td>
<td>St. 4-6/4</td>
<td>St. 16/10d</td>
<td>27-29d</td>
<td>Wing lifted off blastoderm, leg primordium is condensation of mesoderm</td>
</tr>
<tr>
<td>St. 17</td>
<td>29-32 somites</td>
<td>St. 7-8/4</td>
<td>St. 15</td>
<td>34d</td>
<td>Proximal glomeruli first appear after the pronephros has begun to degenerate. The mesonephros starts to differentiate, condensing blastema from somites 13/14-20 limb buds apparent</td>
</tr>
<tr>
<td>St. 18/3d</td>
<td>30-36 somites</td>
<td>St. 9/5</td>
<td>St. 17/10.5d</td>
<td>37d</td>
<td>Anterior, non functional, mesonephric tubules have fused with the mesonephric duct Limb buds enlarged</td>
</tr>
<tr>
<td>St. 19</td>
<td>37-40 somites</td>
<td>St. 12/10.5d</td>
<td>St. 18/11d</td>
<td>41d</td>
<td>Differentiation of principle mesonephric region (somites 20-30) occurring First sign of genital ridge development</td>
</tr>
<tr>
<td>St. 20</td>
<td>40-43 somites</td>
<td>St. 16/11d</td>
<td>St. 19/11.5d</td>
<td>46d</td>
<td>Tip of tail still unsegmented, faint eye pigment</td>
</tr>
<tr>
<td>St. 21</td>
<td>43-44 somites</td>
<td>St. 10</td>
<td>St. 17</td>
<td>51d</td>
<td>Somites extend to tip of tail</td>
</tr>
<tr>
<td>St. 22</td>
<td></td>
<td>St. 11</td>
<td>St. 18-19</td>
<td>5w</td>
<td>Proximal degeneration is complete Ureteric bud appears shortly after the nephric duct opens into the cloaca First sign of Mullerian duct</td>
</tr>
<tr>
<td>St. 23/4d</td>
<td></td>
<td>St. 14</td>
<td>St. 18-19</td>
<td>5w</td>
<td>Appearance of genital tubercle</td>
</tr>
<tr>
<td>St. 26/5d</td>
<td>St. 12-13</td>
<td>St. 14</td>
<td>St. 18-19</td>
<td>5w</td>
<td>Pronephros degeneration is complete Ureteric bud appears shortly after the nephric duct opens into the cloaca First sign of Mullerian duct</td>
</tr>
<tr>
<td>6d</td>
<td>St. 14</td>
<td>St. 18-19</td>
<td>5w</td>
<td></td>
<td>Appearance of genital tubercle</td>
</tr>
<tr>
<td>6.5d</td>
<td>St. 21/22/31-40d</td>
<td>St. 18-19</td>
<td>5w</td>
<td></td>
<td>Appearance of genital tubercle</td>
</tr>
<tr>
<td>7d</td>
<td>St. 22-23/36-45d</td>
<td>St. 18-19</td>
<td>5w</td>
<td></td>
<td>Appearance of genital tubercle</td>
</tr>
<tr>
<td>8-9d</td>
<td></td>
<td>St. 18-19</td>
<td>5w</td>
<td></td>
<td>Appearance of genital tubercle</td>
</tr>
<tr>
<td>10d</td>
<td>St. 24/45-50d</td>
<td>St. 25/16.5d</td>
<td>5w</td>
<td></td>
<td>Appearance of genital tubercle</td>
</tr>
<tr>
<td>11d</td>
<td></td>
<td>St. 25/16.5d</td>
<td>5w</td>
<td></td>
<td>Appearance of genital tubercle</td>
</tr>
<tr>
<td>12d</td>
<td></td>
<td>St. 25/16.5d</td>
<td>5w</td>
<td></td>
<td>Appearance of genital tubercle</td>
</tr>
<tr>
<td>15d</td>
<td></td>
<td>St. 25/16.5d</td>
<td>5w</td>
<td></td>
<td>Appearance of genital tubercle</td>
</tr>
<tr>
<td>15-21d</td>
<td></td>
<td>St. 25/16.5d</td>
<td>5w</td>
<td></td>
<td>Appearance of genital tubercle</td>
</tr>
<tr>
<td>21d approx hatching</td>
<td>St. 28/65d hatching</td>
<td>19d birth</td>
<td>9 months birth</td>
<td></td>
<td>Appearance of genital tubercle</td>
</tr>
</tbody>
</table>

178
major mechanisms of evolution. One of the most obvious changes in the species investigated here is the difference in relative timing of somite and limb bud development. The onset of development of the limb buds in the chick and alligator occurs when a higher number of somites have developed than in mammals. Also the hind limb bud develops first in alligator, as opposed to the forelimb bud in mammals. These two characteristics are used for staging mammalian, chick and alligator embryos and so discrepancies can arise depending on which features of the embryos are compared.

If the developmental events in the genitourinary system are compared against other developmental events, mammalian pronephros differentiation and degeneration occurs at about the same time as in the chick, but mesonephros differentiation and degeneration occurs earlier. Mammalian mesonephros function has never been demonstrated, i.e. it is vestigial. This is in contrast to the chick and alligator where the mesonephros does function and develops as a much larger organ than that observed in mammals (Romanoff 1960).

6.3.2 WTI expression in chick embryos

Chick embryos were harvested between stages 15-20 (53 hrs-3.5 days) (Hamburger and Hamilton 1951) approximately equivalent to 9.5-11 day mouse (Table 6). These stages corresponded to the development of the mesonephros from the urogenital ridge and therefore with expression of WTI observed in mammals (Armstrong et al., 1992). The analysis of WTI expression in chick embryos was prevented in the anterior region due to trapping of the probe. It was decided that the signal present in the anterior region but not the posterior region was non-specific because signal was also obtained with a vector only probe in the anterior but not the posterior region (Figure 6.3c). In early stages WTI expression was observed in the urogenital ridge region of the intermediate mesoderm, extending caudally from the level of the heart, and in later stages expression in the developing paired mesonephroi was seen. Figure 6.3a shows a stage 16 embryo probed with antisense WTI RNA. The paired regions of expression extend caudally from the heart along the ventral surface. At this stage no tubules were visible in this region and the mesonephric duct which induces their differentiation has just reached the posterior end of the embryo (Figure 6.3b), indicating that this may correspond to the early stage of nephron differentiation following induction by the mesonephric duct. By stage 19 tightly packed tubules are visible in mesonephros (data not shown). In the stage 20 embryo
Figure 6.3 Expression of \textit{WTI} in the developing chick

a Whole-mount \textit{in situ} hybridisation of St. 16 chick embryo with an digoxigenin labelled antisense \textit{WTI} RNA probe. Staining of paired ridges in the posterior end of the embryo corresponding to the differentiating intermediate mesoderm (urogenital ridge) is visible. The curve of the embryo hides the fact that the expression extends as far anterior as the heart. The most posterior end of the embryo was lost during dissection as it had not been completely delineated from the blastoderm (15x magnification).

b Drawing of a St. 15 chick embryo (Balinsky 1981) showing the position of the intermediate mesoderm and mesonephros corresponds exactly with the region of \textit{WTI} expression seen above.

c Whole-mount \textit{in situ} hybridisation, with a vector only probe, of a St. 16 chick. Staining is only seen in the anterior region, due to trapping of the probe in intraembryonic cavities (10x magnification).

d A St. 20 embryo lying on its back to show the expression of \textit{WTI} in the developing right mesonephros. The \textit{WTI} signal does not appear to be uniform along the mesonephros and this may correspond to localisation of high expression levels in the developing nephrons. One of the forelimbs is also visible and no \textit{WTI} expression was detectable (12x magnification).

(h heart, im intermediate mesoderm, fl forelimb bud, m mesonephros, n developing nephrons, v vitelline veins)
im
h
v

right vitelline vein

left vitelline vein

mesonephros

spinal cord

open neural plate

remnant of primitive streak

mesonephros

somites

spinal cord

2nd and 3rd visceral pouches

eye

diencephalon

ear vesicle

cranial ganglion V

cranial ganglion VII

Rhombencephalon

pharyngeal membrane

 Rathke's pouch

brain

heart

stomach

anterior portal
the signal is discontinuous along the ventral edge of the mesonephros, consistent with the localisation of WT1 to the differentiating nephrons, perhaps to the glomeruli which develop after the tubules are seen (Figure 6.3d).

6.3.3 WT1 expression in alligator embryos

The collection of alligator eggs from Louisiana, USA, their incubation and in situ hybridisation was performed by AM Coriat at the University of Manchester as part of a collaborative study on early gonad development in alligators. Alligator embryos at stages 3-14 (3-17days) (Ferguson 1987) were investigated. These are equivalent to chick stages 15-29 (2-6days) (Hamburger and Hamilton 1951). Earlier embryos were not obtained because some development has occurred before oviposition and eggs cannot be immediately collected from nests.

No expression was detectable in alligator at stages 3 and 5 (Figure 6.4 a&b). The onset of WT1 expression in the alligator occurs at stage 6 (Figure 6.4), equivalent to about stage 17 in chick (Table 6.1). The expression is localised to the anterior end of the intermediate mesoderm at the level of the heart (Figure 6.4c) and the roof of the rhomboencephalon (hind brain) (Figure 6c&d). In a number of embryos, signal was observed in the forebrain but this was probably due to trapping of the probe as a similar signal was observed in a stage 6 embryo probed with the sense control (Figure 6.4f). In stages 9-11, the expression of WT1 extends down the intermediate mesoderm corresponding with the wave of differentiation of the urogenital ridge induced by the mesonephric duct (Figure 6.5). Two components of the developing circulatory system seem to express WT1. At the level of the developing forelimb bud, a paired region of expression ventral to the mesonephros is observed, this is thought to correspond to the vitelline veins which overlie the mesonephros and above this there is signal in the developing heart. At stages 10 and 11 the signal extends the length of the intermediate mesoderm and is also found in the fore and hind limb buds (Figure 6.5c-e). As the hand plate forms in stages 13-14 the limb bud expression becomes localised to the presumptive wrist and armpit regions and a region adjacent to this on the body wall (Figure 6.6). On either side of these embryos from the neck region downwards signal is also visible, this is thought to correspond to the lateral plate mesoderm, which is differentiating into the body wall. Sections of the mesonephros of stage 21 embryos show the tubules passing through the comma and S-shaped stages of differentiation (Figure 6.7). In situ hybridisation demonstrates localisation of WT1 expression to the tubules.
Figure 6.4 Whole-mount *in situ* hybridisation of alligator embryos - onset of *WTI* expression

a & b St. 3 and 5 embryos probed with antisense *WTI* riboprobe, no expression of *WTI* is detectable.

c St. 6 embryo hybridised with antisense probe. The onset of *WTI* transcription is detected in anterior end of intermediate mesoderm beginning at the level of the heart.

d. St. 6 showing that staining is also apparent in the roof of the rhomboencephalon, the hind brain.

e. St. 7 embryo showing persistence of the hind brain expression

f. St. 6 embryo probed with sense probe. Only non-*WTI* specific signal was found in the fore brain.

(fb fore brain, hb hind brain, h heart, im intermediate mesoderm)
Figure 6.5 WTI expression in stage 9-11 alligator embryos

a &b St. 9 embryos, left and right sides respectively, showing the extension caudally of WTI expression in the differentiating urogenital ridge as far as the cloaca. Expression in the heart and vitelline veins is apparent as is expression in the roof of the hind brain. Also at this stage the otocyst is stained, this is probably due to the trapping of the probe as well as in the forebrain.

c St. 10 embryo, expression of WTI is detected in the limb buds as well as the intermediate mesoderm.

d&e St. 11 embryos either lightly or heavily stained. Expression is maintained in the intermediate mesoderm, limb buds, heart and vitelline veins. (fl forelimb bud, hl hind limb bud, h heart, hb hind brain, im intermediate mesoderm, o otocyst, v vitelline vein)
**Figure 6.6 WTI in stage 13-14 alligator embryos**

a-c. Stages 13-14. The abdominal body wall has formed so the mesonephros is no longer visible. *WTI* expression in the limbs has become to limited specific regions in the forelimbs.

The body wall is also expressing *WTI*, from the neck region down. The expression seen in the heart is probably due to a high level of expression in the mesothelium.

(bw body wall, fl fore limb, h heart)
6.3.4 Xenopus whole-mount hybridisation

In the Xenopus embryos probed with an antisense XWT1 riboprobe (transcribed from PCR clone XWTI) no convincing signal was detected, compared with the sense control. This could be because the probe that was used was too short. It was only 277bp long and the minimum recommended length is 500bp. The isolation of a larger region of XWTI had been attempted by library screening but no clones were obtained. Alternatively the stages used may have been too early. This probably not the case as in the oldest embryos used (stage 39) pronephric tubules were apparent and the nephric duct had extended along the length of the embryo (Lynch and Fraser 1990) and so the induction of the urogenital ridge would be expected to be underway and therefore WTI should be expressed. The dissection of WTI expression in Xenopus will have to await the isolation of a longer probe.

6.4 Comparison of WTI expression

In contrast to the detailed information about the chick genitourinary system, information on the timing of the onset of alligator nephrogenic differentiation is limited. The earliest reports observe the presence of what is thought to be the nephric duct rudiment in stage 1 embryos. In stage 3 embryos 5-6 nephric tubules are present at the level of the heart and this is therefore thought to be the onset of pronephric development (Reese 1908). This corresponds with the timing of pronephric development observed in chick. By stage 5, tubule formation has extended as far as the liver (Reese 1908). The metanephric anlagen has been observed at >40 somites (St. 12) (Forbes 1940) and many metanephric tubules are present by stage 20 (Reese 1908).

In the chick, expression was seen all along the urogenital ridge at stage 16. This is just before the first mesonephric differentiation is observed (stage 17), but after the mesonephric duct has reached the cloaca (stage 15) and therefore at the stage of induction of the mesonephric blastema (Romanoff 1960). The expression in chick just precedes the first observation of WTI expression in the anterior end of the urogenital ridge in the alligator. The stage 6 alligator in which this was seen was approximately equivalent to a stage 17 chick, and so WTI expression in the alligator is detectable around the times when pronephric and mesonephric differentiation is occurring in the chick, but with a slightly later onset than seen in the chick. Expression was detected in the region where the first tubules are observed in the alligator (somites 4-10), but after their differentiation has been reported. It maybe that in the pronephros, where
the nephrons have external glomeruli \textit{WTI} is regulated in a slightly different manner, or that \textit{WTI} levels were below the detection threshold. In both species \textit{WTI} expression extends the length of the urogenital ridge from the level of the heart (similarly to mouse) as far as the cloaca. In both species evidence of localisation to the developing nephrons has been found. These results agree with the timing and localisation of \textit{WTI} expression in the mouse mesonephros (Armstrong \textit{et al.}, 1992).

Several other sites of \textit{WTI} expression were observed in alligator embryos. The possible neural expression is of interest. The expression of \textit{WTI} in the roof of the developing hind brain has not been detected in any species apart from the alligator. Expression in the newborn rat (Sharma \textit{et al.}, 1992), and in 15 day mouse brain and in neural tissue of the spinal cord of the developing human, mouse and rat (Pritchard-Jones \textit{et al.}, 1990, Armstrong \textit{et al.}, 1992, Sharma \textit{et al.}, 1992) has been reported indicating that \textit{WTI} can be expressed in tissues which are ectodermally derived in addition to those of mesodermal origin.

Expression is also present in the lateral plate mesoderm which will give rise to the connective tissue, blood vessels and musculature of the body wall. A transient localisation of signal in the developing striated musculature of the body wall has been observed in day 13 mice (Armstrong \textit{et al.}, 1992). This is a particularly interesting property considering that ectopic striated muscle is histologically observed in Wilms' tumours whereas only smooth muscle is present in normal kidneys (Mierau \textit{et al.}, 1987). The alligators in which this signal was observed were of a similar developmental stage to the mice indicating a possible conserved role for \textit{WTI} in the developing body wall. The alligator expression in the limb buds and the vitelline veins represent new sites of \textit{WTI} expression. The observed heart expression may be analogous to the mouse and human pattern where \textit{WTI} is expressed in the overlying mesothelium which forms part of the pericardium (Pritchard-Jones \textit{et al.}, 1990).
Chapter 7

Discussion
The evolutionary history of \textit{WT1} in the vertebrate classes has been illuminated by the findings presented above.

- \textit{WTI} orthologs have been isolated by a combination of PCR and library screening from representatives of all the vertebrate classes.
- A high degree of conservation, attributable to purifying selection, has been observed between the predicted protein sequences.
- Conservation is particularly stringent in the zinc fingers which are almost identical in all vertebrates.
- The human genomic structure which is one exon for each zinc finger domain has been conserved in zebrafish, including the presence in this species too of the alternative splice donor sequence that is responsible for the KTS insertion between ZF3 and ZF4.
- However two homopolymer domains in the transregulatory domain, the 17 amino acid alternative splice, and the N-terminal region of exon 6 are not highly conserved.
- All residues shown to be mutated in Wilms' tumours, both somatic and constitutional, are conserved.
- The \textit{WTI} expression pattern has been shown to be conserved in the developing genitourinary ridge, in particular the developing nephric tubules.
- Expression of \textit{WTI} outside of the intermediate mesoderm derived tissues has been found in the alligator as well as in mammals.

This study has made a significant start to the comparative analysis of \textit{WTI} structure and function in the vertebrates.

7.1 The evolution of transcription factors in vertebrates

Comparison of the predicted amino acid sequences has revealed very few substitutions in the \textit{WTI} orthologs, particularly in the zinc fingers. The conservation of other transcription factors for which a number of orthologs have been isolated was compared with \textit{WTI}. Table 7.1 shows the percentage conservation for a selection of transcription factors within the vertebrate lineage. The figures were either taken from the quoted paper or calculated using the gap program of GCG from sequences in the GenBank database. The figures are similar to those obtained for \textit{WTI} (Table 4.3 & 4.6), and with all of these there is a striking increase in homology within the
<table>
<thead>
<tr>
<th>Gene</th>
<th>Species Compared</th>
<th>% Conserved Amino Acids in the Entire protein</th>
<th>Region of increased conservation</th>
<th>% Conserved Amino Acids in this Region</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>p53</td>
<td>Human-Mouse</td>
<td>85.5%</td>
<td>Regions 4 &amp; 5</td>
<td>96%</td>
<td>Soussi et al., 1990</td>
</tr>
<tr>
<td></td>
<td>Human-Chick</td>
<td>69%</td>
<td></td>
<td>92%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human-Xenopus</td>
<td>51%</td>
<td></td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse-Xenopus</td>
<td>57%</td>
<td></td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chick-Xenopus</td>
<td>69%</td>
<td></td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Krox-20</td>
<td>Human-Mouse</td>
<td>90%</td>
<td>Zinc fingers</td>
<td>100%</td>
<td>Oxtoby and Jowett 1993</td>
</tr>
<tr>
<td></td>
<td>Human-Zebrafish</td>
<td>81.5%</td>
<td>Zinc fingers</td>
<td>98.7%</td>
<td></td>
</tr>
<tr>
<td>c-myc</td>
<td>Human-Marmoset</td>
<td>96%</td>
<td>Basic Helix loop helix</td>
<td>98%</td>
<td>Whitfield et al., 1993</td>
</tr>
<tr>
<td></td>
<td>Human-Mouse</td>
<td>94%</td>
<td>Basic Helix loop helix</td>
<td>94%</td>
<td>Schreiber-Agus et al., 1993</td>
</tr>
<tr>
<td></td>
<td>Human-Zebrafish</td>
<td>61%</td>
<td>Basic Helix loop helix</td>
<td>81%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse-Zebrafish</td>
<td>63%</td>
<td>Basic Helix loop helix</td>
<td>83%</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Xenopus</em> - Zebrafish</td>
<td>65%</td>
<td>Basic Helix loop helix</td>
<td>83%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trout-Zebrafish</td>
<td>77%</td>
<td>Basic Helix loop helix</td>
<td>98%</td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>Human-Mouse</td>
<td>98%</td>
<td>Basic Helix loop helix</td>
<td>100%</td>
<td>Whitfield et al., 1993</td>
</tr>
<tr>
<td></td>
<td>Human-Zebrafish</td>
<td>86%</td>
<td>Basic Helix loop helix</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>SRY</td>
<td>Human-Marmoset</td>
<td>61%</td>
<td>HMG box</td>
<td>95%</td>
<td>Whitfield et al., 1993</td>
</tr>
<tr>
<td></td>
<td>Human-Mouse</td>
<td>N/D</td>
<td>HMG box</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>SOX3</td>
<td>Human-Mouse</td>
<td>94%</td>
<td>HMG box</td>
<td>&gt;94%</td>
<td>Schult-Merker et al., 1992</td>
</tr>
<tr>
<td>Brachyury (T)</td>
<td>Mouse-Zebrafish</td>
<td>69.7%</td>
<td>DNA binding region</td>
<td>84.5%</td>
<td>Hammerschmidt and Nusslein-Volhard 1993</td>
</tr>
<tr>
<td></td>
<td><em>Xenopus</em> - Zebrafish</td>
<td>76%</td>
<td>DNA binding region</td>
<td>94%</td>
<td></td>
</tr>
<tr>
<td>Snail</td>
<td>Mouse-Xenopus</td>
<td>76%</td>
<td>Zinc fingers 2-5</td>
<td>86%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse-Zebrafish</td>
<td>76%</td>
<td>Zinc fingers 2-5</td>
<td>83%</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Xenopus</em> - Zebrafish</td>
<td>76%</td>
<td>Zinc fingers 2-5</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zebrafish-Drosophila</td>
<td>76%</td>
<td>Zinc fingers 2-5</td>
<td>71%</td>
<td></td>
</tr>
<tr>
<td>Pax 6</td>
<td>Human-Mouse</td>
<td>&lt;99%</td>
<td>paired domain</td>
<td>100%</td>
<td>Ton et al., 1992</td>
</tr>
<tr>
<td></td>
<td>Human-Quail</td>
<td>98.6%</td>
<td>paired domain</td>
<td>100%</td>
<td>Martin et al., 1992</td>
</tr>
<tr>
<td></td>
<td>Human-Zebrafish</td>
<td>96%</td>
<td>homeodomain</td>
<td>100%</td>
<td>Krauss et al., 1991</td>
</tr>
<tr>
<td></td>
<td>Human-Zebrafish</td>
<td>96%</td>
<td>paired domain</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quail-Zebrafish</td>
<td>96%</td>
<td>paired domain</td>
<td>98.2%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quail-Zebrafish</td>
<td>96%</td>
<td>homeodomain</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>
presumed DNA binding domains as compared to the rest of the protein. For both tumour suppressor genes (WT1 and p53) the pathogenic mutations found have been almost all located in these conserved regions (Soussi et al., 1990). This confirms the supposition that positions which have been found to be mutated in pathological states in humans are conserved in evolution. A high level of conservation has been observed for many transcription factors. Changes in the regulatory genes in development could be under a high level of selection because changes in these may affect more than one pathway so it may be a rare event that changes one of these proteins so that there is no deleterious effect on any of the pathways.

7.2 Rate of evolution of WT1

The comparison of the rate of change at synonymous and nonsynonymous sites demonstrated that the former was far greater than the latter, indicating that the slow rate of change of the protein sequence resulted not from an intrinsically slow rate of nucleotide mutation but was due to selection. By comparing the amount of sequence divergence with divergence time a linear trend was seen. The major anomaly seen was that the rate of change in the reptilian lineage seems to be lower than expected. It is difficult to make definite conclusions about rates in different lineages because of the considerable uncertainty about the true divergence times. Differences in the rate of evolution have been seen in a number of taxa, and it has been observed before that the rate of sequence change in the aves is lower than in some other lineages (Britten 1986). Attempts have been made to account for these differences in the framework of the theory of a molecular clock (Wilson et al., 1987). The idea that local clocks exist as opposed to a single global clock, seems to have gained greater acceptance. There is probably a range of variables involved in the rate of sequence change, including generation time effect (Kohne 1970), DNA repair (Britten 1986), or metabolic rate (Martin and Palumbi 1993).

7.3 The transregulatory domain

The mechanism by which sequence specific transcription factors, such as WT1, act to activate or repress expression from their target genes is just beginning to be elucidated. There seems to be a common theme for the diverse regulators, in that they affect transcription from a wide variety of promoters and they can function synergistically and co-operatively with other factors. The stage at which these factors are now thought to act is at the initiation of transcription, by modulating the
formation of the complex between the RNA Polymerase II complex and its associated
general transcription factors, which is now known to be a highly ordered multi-step
process (Zawel and Reinberg 1993). Although the DNA binding regions in these
transcription factors show a limited number of recognisable motifs e.g. zinc fingers,
homeodomains, and bZip motifs, the only similarities between putative transregulatory
domains is a richness in certain sorts of amino acids; the acidic domains or proline or
 glutamine rich (Mitchell and Tjian 1989). Experiments to define the functionally
important regions within these domains have been tried. In several transcription
factors different regions have been found to be acting independently to activate or
repress transcription (Russo et al., 1993, Chi and Carey 1993). In most experiments
the effect of WTI has been to repress transcription from reporter genes and it seems
that disturbing the structure of the transregulatory region of WTI always disturbs the
repression effect of the protein (Madden et al., 1991, Drummond et al., 1992,
had been hoped that the pattern of conservation of this region of WTI would provide
insight into the architecture of this region. No strikingly modular pattern of
conserved and non-conserved residues was observed for WTI which would point to
functional and non-functional regions. In some ways the high level of conservation of
this region was surprising considering only a very small percentage of mutations occur
in this region. When comparing mutated regions with evolutionarily conserved ones,
evolutionary conservation probably provides a better picture of functional regions, as
mutations will only be observed if the phenotypic change is large enough to be noticed
but is not lethal. In addition only a subset of genitourinary abnormalities, mostly
Wilms' tumours, have been analysed for WTI mutations. As yet only somatic
mutations have been observed in the transregulatory domain and it may be that a
constitutional change in this region of WTI may produce a different phenotype to
constitutional zinc finger mutations.

The major difference observed was in the two homopolymer domains of
polyproline (Proline)\textsubscript{13} and polyglycine (Glycine)\textsubscript{5}. The conserved, higher than
average levels, of proline, glutamine, and serine are consistent with a non random
distribution of amino acids found in transcription factor transactivation domains. Also
using the Chou and Fasman program for prediction plots of \(\alpha\) helix, \(\beta\) sheet and \(\beta\)
turns and the Kyte and Doolittle program for estimation of hydrophobicity no great
changes between the different species was shown (Appendix B). The high level of
conservation of the WTI orthologs points to the functional importance of this region
suggesting that WTI is not merely acting as a structural antagonist of \(EGR1\)
(effectively blocking EGR1 binding to its target promoters), and repressing growth in that manner.

7.3.1 Evolution of homopolymer domains

The mechanism by which the difference in these domains arose can be speculated on. (Proline)_{13} is not present in the chick WTI ortholog and the region encoding (Glycine)_{5} has not been conserved. It is not known whether or not the (Proline)_{13} region has been inserted or deleted in WTI, since the mammalian/reptilian divergence. To answer this question will require the isolation of this region from more diverged species. Insertions or deletions can arise from slippage of the DNA polymerase during replication resulting either in replication of a region already replicated or skipping of a region. Slippage is known to happen in the regions of DNA containing contiguous short repeats due to the increased chance in mispairing between these repeats (Li and Graur 1991, Richards and Sutherland 1994). It could be speculated that the two regions (Glycine)_{5} and (Proline)_{13} in WTI have been produced by insertion or expansion of a trinucleotide repeat of CCN and GGN which encode proline and glycine respectively. (Glycine)_{5} may have been produced prior to mammalian/reptilian divergence but was only conserved in the mammalian lineage. Both of these homopolymer motifs are conserved in three mammalian orders, primate, rodent and artiodactyls (Figure 5.3), although the number of residues is not completely stable, indicating that some replication slippage is occurring.

There is growing interest in these homopolymer motifs and their origin. This has been aroused by a new sort of pathogenic mutation, the expansion of trinucleotide repeats, called dynamic mutations (Richards and Sutherland 1992). The first incidence was discovered in the FMR gene. It was found that an increase in the number of CCG repeats in the gene was associated with Fragile X syndrome. Carriers are found to have a greater than normal number of repeats, and in affected individuals this is even further increased. The number of disease genes where this has been found to occur is expanding, e.g. X-Linked spinal and bulbar muscular atrophy (La Spada et al., 1991), myotonic dystrophy (Brook et al., 1992, Fu et al., 1992, Mahadevan et al., 1992), FRAXE site (Knight et al., 1993) and Huntington's Chorea (The Huntington's Disease Collaborative Group 1993). The position within the gene is not the same for all diseases; expanded repeats are found in coding or non coding regions, suggesting that the expansions can interfere with gene expression or function in a variety of ways. Only two trinucleotides have been found to be expanded in this way, CCG and CAG;
Table 7.2 Genes containing homopolymer domains

<table>
<thead>
<tr>
<th>Gene</th>
<th>Homopolymer Domain</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGR2</td>
<td>Proline</td>
<td>Joseph et al., 1988</td>
</tr>
<tr>
<td>Androgen receptor</td>
<td>Glutamine</td>
<td>Tilley et al., 1989</td>
</tr>
<tr>
<td>Yeast ADR6</td>
<td>Asparagine, Threonine</td>
<td>O'Hara et al., 1989</td>
</tr>
<tr>
<td>Human calcineurin A</td>
<td>Proline</td>
<td>Guerini and Klee 1989</td>
</tr>
<tr>
<td>Glucocorticoid</td>
<td>Glutamine</td>
<td>Evans 1989</td>
</tr>
<tr>
<td>Receptor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EBNA2</td>
<td>Proline</td>
<td>Horvath and Schwbach 1993</td>
</tr>
<tr>
<td>Mouse Sry</td>
<td>Glutamine</td>
<td>Gubbay et al., 1990</td>
</tr>
<tr>
<td>GAL11</td>
<td>Glutamine</td>
<td>Suzuki et al. 1988</td>
</tr>
<tr>
<td>MCM1</td>
<td>Glutamine</td>
<td>Passmore et al. 1988</td>
</tr>
</tbody>
</table>
it is not known whether this is because these trinucleotides are more susceptible to expansion or whether it is due to the protein products produced. The mechanism by which these expansions occur has been elucidated in one gene: spinocerebella ataxia type 1. In this gene the normal repeat has a single base change interrupting it; if this is mutated back to perfect repeat it becomes unstable (Chung et al., 1993). Instability of di, tri, and tetraneucleotide repeats has now been seen in some cancers arising from mutations in the DNA repair mechanism (Richards and Sutherland 1994).

On searching databases many other genes have homopolymer domains which may also be attributable to the insertion or expansion of a trinucleotide repeat during evolution but in a much more minor way than that seen in disease (Table 7.2). As yet no expansions in WT1 have been found but it may be speculated that if the region encoding the polyproline region became unstable, expansion could be a mechanism of gene inactivation.

7.3.2 Function of homopolymers

Homopolymer domains are thought to play a role in transcription regulation, since proline and glutamine homopolymers are mostly found in transcription factors. Fusion of these to GAL4 DNA binding domains show that they can change the level of transcription from a reporter gene in vitro (Gerber et al., 1994). The deletion of (Proline)_{13} and (Glycine)_{5} homopolymer domains from the human WT1 in vitro caused only a small reduction in the repression of transcription from reporter genes (Madden et al., 1993, Wang et al., 1993a). Although (Proline)_{13} is interrupted at two positions by alanine residues, these are related to proline by only a single change in the nucleotide sequence i.e. CCN to GCN. This may be due to a selection for a functional domain as proline homopolymer domains have their maximum effect in vitro at 10 residues long (Gerber et al., 1994).

From the Chou and Fasman structure prediction plots no structure is predicted for (Proline)_{13} and (Proline)_{4} (Appendix B). Studies on polyproline in solution have found that two structures, type I and type II helices can be formed, dependent on whether the α carbon is in the cis or trans conformation (Mandelkern 1967). Therefore in WT1 these sequences may be forming this rigid rod type of structure if exposed to the solvent on the surface of the protein. The observation that WT1 protein can exist with or without (in the chick) the (Proline)_{13} sequence but without any other large sequence changes suggests that there is a certain degree of structural flexibility in this region. This would suggest a surface location, as seen for this type of
helix in other globular proteins (Adzhubei and Sternberg 1993). Therefore in mammals it may be involved in the type of protein-protein interactions seen with other proline rich sequences e.g. SH3 target binding sites, which have been shown to have a proline type II helix (Yu et al., 1994). The function of (Glycine)₅ is unknown. Glycine having no side chain is the residue with the greatest propensity for rotation and therefore this region may be expected to be flexible, which is consistent with the prediction of a high level of probability of a β loop in this region (Appendix B).

Why these regions have been conserved in mammals but not in birds may only be revealed by a comparison of the in vivo interactions made by the transregulatory domain with other proteins. It could be that there are other changes in the proteins with which it associates which compensate for the WTJ differences. Another possibility is that the insertion or deletion event which produced the difference in (Proline)₁₃ is not as deleterious as substitutions occurring in this region so the region can be wholly conserved or not at all.

7.3.3 Conservation of the putative leucine zipper

At the C-terminal end of the transregulatory domain there is a putative leucine zipper. This has been proposed to be the means by which WTJ protein molecules could dimerise. It is of particular interest because it is thought that the DDS phenotype arises as a result of a dominant-negative mutation, with the mutant WTJ possibly interacting with wild type WTJ through this domain. The conservation and lack of helix breaking prolines in all species would support the existence of leucine zipper. In conflict with this, structure prediction algorithms do not predict the presence of a coiled coil (leucine zipper) in this region. Therefore it may be that interaction of WTJ protein molecules, leading to a dominant-negative phenotype, occurs in a different manner.

7.4 A region of WTJ with species-specific function?

The most variable region between the orthologs is located in exon 6, immediately adjacent to a perfectly conserved region. Comparison of the chick and alligator orthologs reveals throughout the protein that the same positions are often substituted when compared with mammalian WTJ. N-terminal to exon 6, the amino acids found at these substituted positions are often different in chick and alligator, which points to flexibility of the sequence in this region. In the highly substituted
region of exon 6 substituted amino acids are almost identical in chick and alligator. This suggests that some selection may be acting on this region which is different in the two lineages i.e. the sequence difference did not arise from sequence drift due to lack of function but a change in functional properties. This region is adjacent to the 17aa alternative splice (exon 5) which has not been detected in chick or alligator. This region is thought to have a function in the human for several reasons. A mutation in this region was found in a benign mesothelioma and is associated with a change in transrepression activity (Park et al., 1993a). In the same Wilms' tumour cell line that excluded exon 2 from transcripts, exon 5 was also not included, suggesting a role for exon 5 in tumorigenesis (Haber et al., 1993). Also the inclusion of exon 5, converts the +KTS isoform of WT1 from a weak repressor into a strong repressor of WT1 transcription in vitro (Rupprecht et al., 1994).

Species differences in regulatory genes are very important in development as they may be the source of species-specific features. The genitourinary system in which WT1 is expressed has many specialisations within the vertebrates; the most obvious of these being the development of the metanephros. In addition there are differences between classes in the organisation of the genital and urinary ducts. In particular the female genital ducts become highly specialised for viviparity and oviparity. Therefore the changes in exons 5 and 6 may give rise to subtle changes in WT1 activity, which then plays a part in developmental variations in the intermediate mesoderm.

7.5 Selection of alternative splice sites

The conservation of the KTS alternative splice agrees with its importance as suggested by human mutation analysis. The importance of the + and - KTS isoforms being present at the correct ratio is underlined by the observation of two independent heterozygous mutations at the alternative splice site in DDS patients which completely inhibits the inclusion of this sequence. The perfect conservation of the zebrafish intron in the region adjacent to the alternative splice donor sites at the end of exon 9 which results in the inclusion of the KTS sequence, points to the context of a sequence influencing splice site choice. This seems to apply even when the two sites are only 9 bases apart, significantly this is almost one turn of the B-DNA helix. The mechanism by which alternative splicing occurs is unknown but it is thought to be closely related to constitutive splicing, the main elements of which are the snRNPs (small nuclear ribonucleoprotein particles) and the SR proteins (these contain
alternating serine and arginine residues in the C-terminal domain). There are several evolutionarily conserved SR proteins, although the presence of only one seems to allow splicing to occur. Therefore the recruitment of different SR proteins may affect the splice site chosen (Zahler et al., 1993). It may be that the sequence of intron 9 is conserved to favour the binding of particular SR proteins. The situation is complicated as there seem to be multiple ways by which a transcript can be alternatively spliced (McKeown 1992).

The conservation of the KTS alternative splice is consistent with information from other genes. Alternative splicing is found in many developmentally important transcription factors, in PAX6 the same alternative splices have been detected in human as in quail (D. Englekamp personal communication) and the alternative splice 5a which inserts 14aa in the paired domain is identical in mouse and zebrafish (Glaser et al., 1992). In max (a protein that dimerises with c-myc) the same alternative splice is seen in zebrafish and in human (Schreiber-Agus et al., 1993).

To try to understand why exon 5 was not present in non mammalian transcripts the splice sites in the cDNAs were examined. Although most of the splice site consensus occurs in the intron there is a bias in the first two positions in the exon. Some of the least typical splice sites occur around exon 5. Non-consensus splice sites have been implicated in alternative splicing (Jackson 1991). Therefore these differences may be the reason why exon 5 is excluded from transcripts in chick and alligator. The isolation of the relevant genomic sequences will be able to resolve whether the sequence of exon 5 is actually conserved outside of mammals.

RNA editing is thought to be functionally associated with splicing as in vitro cross species editing sometimes results in correct editing but incorrect splicing. It may be that the differences in the editing of exon 6 (Section 5.2.7) and the splicing of exon 5 i.e. the 6bp insertion in the chick and alligator at the editing position in mammals and the lack of a +17aa isoform in the reptilian lineage, is in some way connected. The editing in rat was found to occur in a developmentally regulated manner and was first observed after birth when kidney development is virtually complete. RNA obtained for chick and alligator was only obtained from embryonic sources so it may be interesting to look at later stages to see if any editing could be detected.
7.6 Evolution of the zinc fingers

The data obtained here show that the rate of evolution of the zinc fingers of WT1 is slow, and slower than the rest of the protein (Table 5.5). This is in agreement with evolutionary analysis of other zinc finger genes. The comparisons of Krox 20 and Krox 24 (EGR2&EGR1) (human to fish) (Lanfear et al., 1991, Oxtoby and Jowett 1993); snail (mammals to Drosophila) (Sommer et al., 1992, Hammerschmidt and Nusslein-Volhard 1993); ZFY family (mammals to birds) (Lanfear and Holland 1991), Kruppel and hunchback (arthropods) (Sommer et al., 1992) show a pattern of substitution very similar to that obtained for WT1. The conservation between runs of fingers and homologous fingers has been analysed statistically (Jacobs 1992) and the substituted positions closely resembled the ones found in this study i.e. mostly between the two cysteines or the two histidines or after the second cysteine but not at the 'tip' of the finger. The pattern of substitution would suggest that the WT1 orthologs retain the same DNA binding consensus sequence as human WT1. Although variability is limited to regions of no known function, other positions for which no function is known were also found to be invariant. At these positions mutations have not been observed and they are not conserved in the EGR fingers. The lack of mutations of these sites could arise from the rarity of WT1 mutations, most occurring at CpG dinucleotides. Mutations at the invariant sites could be deleterious enough to be selected against but do not produce abnormalities that have been investigated. The fact that these residues are not present in the EGR zinc fingers but are conserved in WT1 would suggest that these residues are acting to modify the properties of the fingers thereby making the DNA sequence specificity of WT1 subtly different from that of EGR1 and 2.

Families of similar zinc fingers are now being recognised such as the Kox and Fax families (Nietfield et al., 1993, Tunnacliffe et al., 1993). The latter have similar zinc fingers but also contain the Fax motif. They are found in clusters in the vertebrate genome indicating that they have probably arisen by duplication of an ancestral gene. WT1 belongs to the EGR family by virtue of the similarity of its fingers; this family includes EGR1 (Sukhatme et al., 1988), EGR2 (Joseph et al., 1988), EGR3 (Patwardhan et al., 1991), and EGR4 (Crosby et al., 1992). These have three fingers, ZF1 and ZF3 having a very similar consensus sequence in the DNA binding region (type 1) while ZF2 is slightly different (type 2) (Figure 1.2c). WT1 fingers 2-4 are very similar to the EGR fingers in the DNA binding region but WT1 has an extra zinc finger (type 3) which is most similar to the first finger in the Spi family. The Spi family are a set of genes which also have three fingers, similar to the
first three fingers of WTI. This includes Sp1 and Sp2 (Kadonaga et al., 1987, Kingsley and Winoto 1992), SPR-1 and SPR-2 (Hagen et al., 1992), Drosophila buttonhead (Wimmer et al., 1993), and rat BTEB (Imataka et al., 1992). The EGR-like second and third fingers are slightly different in that the first predicted backbone contacting residue is a lysine not an arginine, but this is a conservative substitution of one amino acid with a basic side chain for another. The type 1 finger is also found in two yeast genes, MIG1 in Saccharomyces cerevisiae (Nehlin and Barnes 1990) and Cre-A in Aspergillus nidulans (Dowzer and Kelly 1991). Both genes are involved in the regulation of catabolite metabolism and have two zinc fingers in tandem, the second one resembling the type 1 finger.

The evolutionary relationship between these fingers is unknown; it is not clear whether it is due to homology arising from a common ancestral gene or whether the similarity has arisen as a result of adaptive changes producing convergent evolution for GC rich DNA binding sites. The most striking resemblance is not the homology of the individual fingers but the way in which the different fingers are in a similar order. Amongst these genes, for which the genomic structure is known, WTI is unique in having each finger in a separate exon as opposed to the others where the fingers are in a single exon. The intron/exon junctions occur at the same position in each finger of WTI and in the same phase, specifically within the glycine of the linker region. This is a similar genomic structure to that of TFIIIA in Xenopus (Miller et al., 1985, Tso et al., 1986), where all nine fingers are in separate exons with neighbouring fingers having similar exon junctions. From this arrangement it has been proposed that these fingers arose by duplications of an initial exon, as seen in other independent structural motifs such as kringle and growth factor domains (Patthy 1985). The process by which these modules of structure are duplicated and moved around is known as exon shuffling (Doolittle 1981). It has also been proposed that these genes have given rise to other genes with multiple fingers per exon (El-Baradi and Pieler 1991). This could have occurred by loss of introns or by reverse transcription of the mRNA transcripts and integration into the genome. Reverse transcription is a not an uncommon event, with the amount of the vertebrate genome produced by reverse transcription events estimated to exceed 10% (Temin 1985).

Therefore is WTI ancestral to the EGR or Sp1 families? No similarities between WTI and the EGR or Sp1 families are found in addition to the homology of the zinc fingers. Whereas within the Sp1 and EGR families further homologous regions have been observed in the transactivation domain (Crosby et al., 1992, Wimmer et al., 1993). Therefore if they are related to WTI by ancestry there has been a complete
divergence of the rest of the gene or the gene was only partially duplicated. For \textit{WTI} to be the ancestral sequence it would have to have evolved far earlier than so far detected as the \textit{Sp1} homolog \textit{buttonhead} has already been described in \textit{Drosophila}. Intense search for, isolation and comparison of more orthologs will answer this question. The total genome sequencing programs that are underway in several species will permit the presence or absence of these genes to be definitively determined and provide information about the origin of these genes.

Convergent evolution may have been a major determinant of the similarity between these genes (Doolittle 1994). This could be a result of the limitations imposed on the sequence of the fingers by the fairly rigid structure holding it together and their interaction with a very invariant structure i.e. DNA. Database searching for \textit{TFIIIA}-like fingers shows that the residues in common between the \textit{WTI}, \textit{EGR}, and \textit{Sp1} fingers are very common in other fingers. This supports the fact that there are only a limited number of options allowing DNA binding. These similarities combined with the number of reported fingers (1340 in 1992, Jacobs 1992) means that many fingers may appear related but are not. Then again, it may be speculated that all fingers are related as they all may have arisen from a single ancestral finger. This ancestral finger probably arose after the divergence of pro- and eukaryotes, as only eukaryotes have been found to contain the \textit{TFIIIA}-like zinc fingers.

7.7 Conservation of expression of the \textit{WTI} gene in the intermediate mesoderm

PCR and Northern blot analysis of pig, chick, alligator and \textit{Xenopus} RNA from the intermediate mesoderm derived organs revealed expression of \textit{WTI}. This demonstrates that in addition to the conservation of \textit{WTI} at the level of protein sequence there is also conservation in the pattern of gene expression. This would indicate that at least part of the function of \textit{WTI} has been conserved in mammalian, reptilian and amphibian lineages. In the case of \textit{Xenopus}, \textit{WTI} expression in whole tadpole and mesonephros were compared. It was found that \textit{WTI} expression was only detectable in the mesonephric RNA. This would suggest that expression is specific to the mesonephros. Unfortunately the samples compared were not from the same stages, so the presence of \textit{WTI} could correspond to a general up-regulation of expression in the embryo, but it is more likely that it is due to localisation of \textit{WTI} in the differentiating kidneys.

In pig, chick and alligator the size of \textit{WTI} transcript was found to be conserved when compared to human and mice. In \textit{Xenopus} the size of the mRNA transcript had
increased to 5kb. Until a more complete *Xenopus WT1* cDNA is cloned it will not be known whether this is due to a change in the coding or non coding regions.

In chick an additional 2kb transcript was also seen; its identity is unknown. Interestingly a smaller 2.5kb testis specific transcript has been detected in mouse and human (Pritchard-Jones *et al.*, 1990, Pelletier *et al.*, 1991c), and also in the developing rat kidney (Sharma *et al.*, 1992). Only the full length RNA was detected in pig testis of an equivalent age to when smaller transcripts are detected in mouse and human. In the chick in the mesonephric RNA alone no second transcript was detected so the small transcript may be gonad specific (Figure 6.1). The origin and functional importance of these smaller transcripts is not known; they may not be essential products in all species and could be transcriptional artefacts. These may result from a high level of transcription initiation with a second cryptic promoter being used or they maybe produced by the use of the stronger polyadenylation signal located in the 3' untranslated region in all species examined (Section 5.2.11).

**7.7.1 Urinary system**

In both the chick and alligator, expression of *WTI* was detected by *in situ* hybridisation in the developing genitourinary ridge and localised to the developing nephrons. This demonstrates that there is a role for *WTI* in nephrogenesis outside of the mammals. Consistent with a role in the earliest differentiation of the nephrons, the expression of *WTI* in chick is observed with the onset of mesonephric differentiation. The conservation of expression suggests that Wilms'-like tumours in other species (such as pig, rat, chick and Japanese eel) could also result from loss of *WTI* function disturbing nephrogenesis. This may now be investigated by looking in these tumours for overexpression of *WTI* and mutations, both of which occur in human tumours. This will permit the proportion of *WTI* mutants found to be compared to the 10% incidence seen in human sporadic tumours. Wilms'-like tumours can be experimentally induced in many vertebrates, so that a greater number of tumours, as compared with the relatively few human cases, can be analysed. Therefore do these treatments affect the *WTI* gene or another WT locus? Another unresolved issue is the relationship of *WTI* with *novC*, the *IGF1* binding protein-like gene implicated in chick virally induced tumours (Joliot *et al.*, 1992). Is this gene in the same pathway as *WTI*, either upstream or downstream?

Wilms' tumours are just one of many kidney diseases (Mierau *et al.*, 1987). The involvement of *WTI* in other kidney malfunctions, apart from the nephropathy seen in
DDS, is not known. The role of WTI in other conditions can now be investigated. As suggested earlier constitutional mutations of WTI in regions apart from the zinc fingers may produce different phenotypes. Rats are thought to be a considerably better model for kidney development and Wilms' tumour, as Wilms' tumours are seen in rats, but not in mice. Currently RWT1 is being used to investigate WTI expression in the postnatal rat, with the specific aim of seeing whether there are changes in WTI expression in pathological states known to be associated with changes in renal growth. Kidney infection in early life (infantile pyelonephritis) which can interfere with renal growth can lead to permanent kidney damage, and is the greatest cause of kidney failure in later life. Preliminary results suggest that WTI expression is affected on infection, so that modulation of WTI activity on infection may help to avoid the permanent damage (E. Ostlund, Karolinska Hospital, Sweden).

7.7.2 Gonads

In stage 20 alligator gonads, just preceding the onset of sexual differentiation (stage 21-22 testis and stage 22-23 ovary), WTI expression was detected by Northern blotting and PCR. Expression appeared to be at approximately the same levels at both male and female producing temperatures. In mammals the indifferent gonad also expresses WTI.

The processes that lead to gonadal sex determination in vertebrates seem to be highly variable: sex can be determined genetically, environmentally or a mixture of both. Mechanisms can vary within classes. In mammals it is the presence of the Y chromosome carrying the SRY gene which leads to the development of the male phenotype. SRY is also known as TDF (testis determining factor) because it is the differentiation of the testis from the indifferent gonad which is important in sex differentiation, and which SRY controls. Therefore is WTI functionally related to SRY? WTI is important for gonad development in both sexes, homozygous deletion of WTI in mice (Kreidberg et al., 1993) results in an inability to form gonads, as opposed to the sex reversal seen when SRY is deleted. WTI is expressed from the earliest stages of genital ridge development i.e. in the 9 dpc mouse (Armstrong et al., 1992) and is followed by SRY expression at 10.5 dpc (Koopman et al., 1990). Putative WTI binding sites have been reported in the SRY gene promoter (Behlke et al., 1993) suggesting that WTI could lie upstream of SRY (Bogan and Page 1994). The localisation of WTI expression in males to the Sertoli cells, which are important for sex determination, further suggests this. WTI however appears not to affect
primary sex determination but it can perturb secondary sex determination. In humans changes in dose (WAGR) or activity (DDS) of \textit{WTI} lead to genitourinary abnormalities; these are likely to be due to disturbances in the development of the gonad which affects hormone production on which secondary sex determination is dependent.

In alligator sex determination is thought to be wholly temperature dependent. In males sexual differentiation of the gonad occurs during the temperature dependent stage, and in females just afterwards (Ferguson and Joanen 1982). Testis determination, similarly to mammals, is thought to be dependent on Sertoli cell differentiation (Ferguson and Joanen 1982, Smith and Joss 1993). In alligators many HMG-box containing genes including \textit{SOX} genes have been identified but not an \textit{SRY} ortholog (Coriat \textit{et al.}, 1993). \textit{SRY} is known to be evolving at a much faster rate than expected when compared with other genes (Tucker and Lundrigan 1993, Whitfield \textit{et al.}, 1993) suggesting that \textit{SRY} may be involved in sex determination only in mammals. Mammalian gonad development is known to differ from other vertebrates. In mammals gonadal sex reversal is very rare, but in other vertebrates this is seen to occur much more frequently e.g. intersexes in amphibians, temperature sensitive reversal of genetic sex in axolotl (Dournon \textit{et al.}, 1990) or by exogenous hormonal disturbances e.g. chicken (Elbrecht and Smith 1992) and alligator (Lance and Bogart 1991&1992). This may suggest that in other vertebrate classes the production of steroid hormones in development may have a greater capacity to influence gonadal sex determination. It is known that steroid hormones can be produced in very early gonadal cells (Haffen 1970). The conservation of \textit{WTI} in alligators and chick makes it likely that its role in early gonad development, and therefore in the hormone producing cells, is conserved. Therefore it is possible, in some species at least, that \textit{WTI} plays a role in sex determination. The \textit{WTI} alligator ortholog is being used as a tool to probe early differentiation, first as a gonadal marker and maybe subsequently to isolate the genes involved in temperature dependent sex determination (AM Coriat, Guy's Hospital, London). One of the other sub-classes in which sex determination is being investigated is the metatheria (marsupials). This is because this sub-class shows most of the features of mammalian (eutherian) sex determination i.e. a testis determining Y chromosome, but also some effects of X chromosome dosage (Graves 1987). Therefore metatheria probably separated from the eutheria as the Y chromosome/\textit{SRY} system of sex determination was evolving. Again in marsupials the predicted protein sequence of \textit{WTI} is highly conserved indicating that the mechanisms that control early gonad development maybe similar to those in eutherian mammals.
7.8 Additional sites of expression

Similarly to mouse and human, sites of WT1 expression in alligator were observed outside of the intermediate mesoderm, in derivatives of the lateral plate mesoderm and the ectoderm i.e. the limbs, body wall, heart, vitelline veins, and hindbrain. In the mouse expression has been observed in the lateral plate mesoderm, specifically the musculature of the body wall, and in defined areas of the developing spinal cord and brain (Armstrong et al., 1992). The expression in the heart is probably a result of the surrounding mesothelium expressing WT1. The vitelline veins contain endothelial lining cells which are a type of epithelium that differentiate from the angioblastic mesoderm, derived from the splanchnic mesoderm. These cells also undergo a mesenchymal-epithelial transition which may explain why WTJ is detected (Balinsky 1981). These cells in mammals have not been observed to express WT1 but, like podocytes, a major site of WT1 expression, they are a subset of epithelia with special permeability properties. These results indicate that WT1 expression may not be limited just to the intermediate mesoderm. The importance of most of these expression sites of is not known. No defects have been reported in the Wt1 knockout mouse, in the limbs or body wall (Kreidberg et al., 1993). This may be because either that there is a species difference, or a redundancy of function in the developmental pathway, or that defects were too subtle to observe, or an effect of genetic background. In all these cases further detailed analysis of WT1 expression in histological sections will or will not confirm the preliminary results above. If expression is seen, detailed localisation will be informative.

7.9 WT1 and limb development

The expression of WT1 in the alligator limbs is of interest because of the known association of limb and genitourinary deformities in several species (e.g. human, mouse and axolotl). This implies that in these species there may be similarities between these systems. The WT1 expression in the limb buds has not been reported before. That it is a bona fide result in alligators is suggested by the subsequent localisation of the expression to presumptive wrist and armpit regions at the time of hand plate formation and its disappearance from the hind limbs.

There is evidence for a connection between kidney and limb development. The limb buds develop in close proximity to the mesonephros, and the mesonephros is known to promote the normal differentiation of chick limb explants in culture, partly due to the production of IGF1 (Geduspan and Solursh 1993). The limb buds are
produced by an initial proliferation of the lateral plate mesoderm beneath the epidermis, cells then condense and transform into chondrocytes to produce the skeletal elements (Gilbert 1991). It is thought that the muscular and dermal elements are contributed the somitic mesoderm (Balinsky 1981). The differential morphogenesis of the fore and hind limbs is thought to arise as a result of the action of regulatory factors in the control of processes common to all four limbs, such as fibronectin accumulation and mesenchyme condensation, but these are differentially regulated in the limbs (Downie and Newman 1994). The expression of WTI in only the fore limb may reflect a role in this.

Several genes have been found to be expressed in both lateral plate and intermediate mesoderm derived structures, specifically in the developing limbs and kidneys. The bone morphogenetic protein BMP-2 is expressed in the mesonephros and the nephric duct (C. Tickle personal communication) as well as in the limb buds, tooth buds, and craniofacial mesoderm (Lyons et al., 1990). The BMP family were originally isolated due to their inductive capabilities in bone formation and belong to the TGF-β superfamily. The TGF-β family is also known to be expressed in the developing kidneys (Hammerman et al., 1992). BMP-2 is now thought to have a wider role than just osteogenesis, being important in mesenchymal-epithelial interactions, like those required for kidney formation (Rosen and Thies 1992). The presence of bone in Wilms' tumours could therefore be due to the nephrogenic cells responding to the presence of BMP-2 when the normal developmental pathway is disrupted.

Ld (limb deformity) mice were produced by the disruption of a gene due to the insertion of a c-myc transgene and have deformed limbs and renal aplasia (Woychik et al., 1990). The gene affected is the formin gene and its expression in mouse and chick has been studied. In the chick, the gene is expressed in the pronephros, in the developing tubules and the glomeruli of the mesonephros (Trumpp et al., 1992). The protein is detectable in the limb bud as it differentiates, and the expression becomes concentrated in the condensing mesenchymal cells destined to form the long bones and digits. In addition to those sites affected by disruption of the gene, expression is seen in the notochord, floor plate and ventral horns of the developing chick neural tube. The latter includes the small region in the spinal cord seen to express WTI in mouse (Armstrong et al., 1992).

There is another mutation whose effect seems to be localised to the genitourinary system and the limbs. This is the short toes (s) mutation in the Mexican axolotl (Ambystoma mexicanum). This homozygote lethal mutation results in renal
aplasia which seems to affect mainly the glomeruli, which are reduced in number or absent. There is also an overall disorganisation of the kidney structure. At the molecular level there is known to be a defect in the disposition of laminin A chain which is randomly distributed in mutants but localised to the basement membrane in wild type kidneys (Washabaugh et al., 1993). Laminin A chain production is thought to play a role in the conversion of the mesenchyme to epithelia in the kidney because it is first detected at the earliest stages of nephron development (Ekblom 1989). The mutation also affects the development and regeneration of the limbs but not of the tail. Differences in laminin A chain deposition have also been observed in mutant limbs (Del Rio-Tsonis et al., 1992). The investigation of a possible role for WTI has been greatly facilitated by the isolation of the axolotl WTI gene (P.A. Tsonis personal communication). Expression patterns in the mutant and wild type are now being examined (K. Del Rio-Tsonis personal communication) and mutational analysis of the gene is also planned.

It may be that the similar set of genes expressed in the limb, kidneys and the CNS reflect a similarity in the developmental decisions and morphological changes being made in the cells. The *crocodilia*, to which alligators belong, closely resemble the primitive archosaurs from which reptiles, birds, and mammals evolved. It may be that WTI expression in the limbs is a feature which has been lost in subsequent evolution of vertebrates. The patterning of the limb bud is due to the ZPA (zone of polarising activity located in the posterior region of the limb bud). This can be mimicked by notochord which is involved in patterning of the mesoderm (Ingham 1994) and is thought to play a role in determination of the urogenital ridge (Etheridge 1969). Therefore WTI expression may be elicited in both the limb and the intermediate mesoderm because the same spectrum of transcriptional regulators are present in both.

### 7.10 Evolution of the vertebrate kidney and WTI

One of the major reasons for isolating the WTI orthologs was to compare WTI functionally and structurally in the context of the evolution of the vertebrate kidney. The two major evolutionary innovations have occurred in the vertebrate kidney. First came the initial evolution of the nephric system, probably pronephros-like, at about the same time that the vertebrates appeared. The second was the formation of the ureteric bud and hence the metanephros at about the same time as the appearance of the reptiles. The mechanisms which underlie these changes are unknown. The
expression of \textit{WTI} in the mesonephros of amphibians demonstrates that the mesonephros of anamniotes and amniotes have similarities at the molecular level. The molecular basis for the degeneration of the mesonephros in amniotes but its retention in anamniotes is also unknown. \textit{WTI} missense mutations are involved in the degeneration of the glomeruli in DDS, therefore it may be that \textit{WTI} is somehow involved in the normal degeneration process. The further investigation of \textit{Xenopus} \textit{WTI} will allow these questions to be answered.

The origin of the vertebrate genitourinary system is not known. In non-vertebrate chordates (i.e. those species with a notochord but no backbone) no nephrons or kidneys have been found; in fact the whole region corresponding to the intermediate mesoderm is not present. In the non-vertebrate chordates other methods of excretion are used such as nephridia (amphioxus) and storage kidneys (ascidia). It is thought that coelomoducts, i.e. mesodermally derived ducts which can lead from the coelom, could be the precursors of the nephrons as these are very similar to the nephrotomes seen in association with external glomeruli. These ducts are observed in many phyla (Barnes \textit{et al.}, 1993). The one coelomoduct found in amphioxus, Hatscheck’s pit, develops in the region, at the border of the lateral plate and segmental (somitic) mesoderm, the location of the intermediate mesoderm in vertebrates (Grove and Newell 1953). Therefore how did the intermediate mesoderm evolve and how did the differentiation of genitourinary system come about? All these structures basically develop from the differentiation of tubules in this region which would suggest that this was an initial development.

The specification of the vertebrate mesoderm is being unravelled. In the anterior-posterior direction the domains of expression of the \textit{Hox} genes is thought to be important. These may be the factors controlling the extent of nephric differentiation and degeneration along the urogenital ridge, as many \textit{Hox} genes are expressed in the kidney and its precursors (Bard and McConnell 1994). The dorsal-ventral axis is best understood in \textit{Xenopus}: there is an initial dorsal-ventral specification where the dorsal vegetal pole induces the notochord and muscle and the ventral vegetal pole induces blood, mesenchyme and mesothelium. There is then thought to be a dorsalising signal (Dale and Slack 1987) which induces the formation of the intermediate mesoderm types, including the pronephros (Lettice and Slack 1993). In the chick, from fate maps, the lateral plate and somatic mesoderm give rise to the intermediate mesoderm (Lear 1994). At the molecular level the genes involved are just beginning to be isolated, i.e. \textit{BMP-4} (Dale \textit{et al.}, 1992) and \textit{Xwnt8} (Christian and Moon 1993) have both been shown to have ventralising activity, and \textit{goosecoid}
has a dose dependent dorsalising activity (Niehrs et al., 1994). In contrast very little is known about axial-paraxial specification. During evolution it would not be hard to imagine that shifting the expression boundaries of one of the genes involved in these specifications may have taken place and so allowed an additional domain to be formed.

With the increase in molecular and morphological evolutionary data, the question of how morphological and molecular evolution are related is able to be addressed. Throughout the vertebrates, the association of WTI with a typically vertebrate structure raises the question of whether WTI arose around the time of vertebrate evolution and was a key to the initiation of kidney evolution. WTI is expressed during the differentiation of the coelomic epithelium into mesothelium, the gonads, and the nephrons; specifically during the transition from mesenchyme to epithelia in these tissues. The chordates belong to the deuterostomes, which are distinct from the protostomes. In the deuterostomes coeloms are derived as invaginations of the gut, in the protostomes coeloms are formed by splitting of the mesoderm (Gilbert 1991). Therefore WTI may be associated with coelomic development in all deuterostomes. The amplification by PCR of a non-vertebrate chordate (ascidian) WTI gene was attempted but was not successful. The information on the vertebrate sequence may be useful in achieving this.

The mesenchymal to epithelial cell type change was originally thought to be unusual (Pritchard-Jones et al., 1990) but it is found in many systems. WTI is known to be involved only in some of the transitions in mouse development (Armstrong et al., 1992). One developmental event in which the transition is being characterised is the differentiation of the mid gut of Drosophila (Tepass and Hartenstein 1994). The endodermal midgut precursors are induced to become epithelial by their interaction with the adjacent visceral mesoderm. This is especially interesting considering the relationship of the deuterostome coelom to the gut and it recently been proposed that there may be conserved elements in gut development in vertebrates due to the similar expression of rat HNF-4 and its Drosophila homolog (Zhong et al., 1993). Therefore if a Drosophila WTI ortholog was sought this may be the best place to start looking. This of course is speculation but recent data to some extent has hinted that evolutionary distant systems may have a similar molecular basis. This in particular has been found in the resemblance in the pattern specification genes, i.e. the expression patterns of the hedgehog genes in vertebrate and to the hedgehog gene in Arthropoda (Ingham 1994). There are now reports of the conservation of gene expression in structures with similar functions but distinct
evolutionary origins e.g. the expression of PAX6 in the vertebrate eye and brain and in
the sensory cells of the anterior region of C. elegans (Chisholm and Horovitz 1994,
Dawkins 1994). This could be a by-product of genes being activated in similar
regions, due to similar patterning mechanisms in these diverged species. These genes
may have been co-opted later into specialised structures in a similar region. Therefore
this raises the question is WT1 involved in the formation a the functional homolog of
the kidneys; the gut derived Malpighian tubules of insects?

7.11 The origin of WT1

In the evolutionary comparison of species and genes the question that arises at
the same time as "why a gene has evolved?" is "how has it evolved?". One of the
major ways that molecular evolution is thought to occur is by gene duplication.
Duplication leads to the release of one copy from selection pressure. In many cases
this will lead to the loss of function of one copy, as evidenced by the presence of
unprocessed pseudogenes, but in other cases before function is lost a separate role for
the second gene may be acquired and subsequently selected for. The increase in
genetic information may then favour the acquisition of new characteristics. The most
well known example is the duplication of the homeobox gene clusters that seems to
have occurred around the time of chordate and vertebrate evolution (Pendleton et al.,
1993). The origin of WT1 is not known. Using current technology there are no
detectable homologies to any other genes outside the zinc fingers. Knowledge of the
tertiary structure of the transregulatory domain may allow identification of structural
homologies when sequence homology is lost, as is now happening with the protein
structures of enzymes (Burley 1994). Only a limited number of protein folds have
been detected in the enzymes suggesting that it is far easier to modify an existing fold
rather than to create a new one. The presence and position of introns between the
zinc fingers of WT1 leads to the suggestion of exon duplication having given rise to
these. If this is the case then there may exist in more diverged species WT1 orthologs
that have 3, 2 or even a single finger. Again the sequence analysis of whole genomes
will allow the presence or absence of WT1 to be unequivocally determined.

7.12 Future work

This project has only just begun the evolutionary analysis of WT1. It can now
be extended at the structural, functional and expression levels. The immediate aims
must be to consolidate this work by isolating complete sequences for the anamniotes,
giving greater depth to the comparison particularly in the transactivation domain and also to define exactly the cell types expressing \textit{WTI} in whole-mount \textit{in situ} hybridisation. The apparent conservation of nephrogenesis and nephroblastoma and the conservation of the \textit{WTI} sequence and expression pattern suggest that study of \textit{WTI} across the vertebrates will be an appropriate method for understanding \textit{WTI} function in development and in Wilms' tumours. There are several projects that have been referred to above, in which the expression of \textit{WTI} in various species is being investigated. In the future it will be particularly desirable to perform whole-mount \textit{in situ} on the zebrafish, as these embryos are transparent and therefore particularly good for investigation of internal structures, especially as then it will be possible to fulfil the aim of looking at the pronephros and external glomeruli. The existence of a \textit{WtI} knockout mouse means that it is possible test whether these orthologs are functional in a different species by rescuing the lethal phenotype or perhaps more interestingly, there may be partial rescue. Also, would the introduction a \textit{WTI} ortholog affect extent of differentiation and degeneration of the urogenital ridge. This would also test whether the unconserved regions are truly non-functional. One of the newest model organisms is the pufferfish (\textit{Fugu}) which is known to have a genome one seventh of the size of the human genome, and genes are thought to cover correspondingly smaller genomic regions. Isolation of whole genomic regions of genes from pufferfish and introduction into other species would be a much simpler task than using orthologous genes from other vertebrates (Brenner \textit{et al.}, 1993). This is now being pursued with the isolation of a \textit{WTI} PCR clone for zinc fingers 1-3, which is three times smaller than the region in human and zebrafish (S. Macrae personal communication).

Finally the \textit{in vitro} characterisation of \textit{WTI} function may be facilitated by comparing the actions of the different orthologs. This may be useful in defining the important regions of the transregulatory domain, especially with reference to the homopolymer domains.
A Final Evolutionary Comparison

Awkward ages

Human: 11-13 yrs.
Dog: 6-8 mos.
Frog: 12-16 wks.
Elephant: 4-6 yrs.
Shark: 1½-2 yrs.
Appendix A
Sequences of WT1 orthologs

The sequences of WT1 orthologs obtained in this study are shown as well as the published human, mouse and rat sequences. Where PCR clones were obtained in addition to library clones the region obtained by PCR is indicated. For the genomic sequence obtained from zebrafish the predicted exons are shown in bold and the predicted cDNA sequence also shown. In the 3’ untranslated regions putative polyadenylation signals are highlighted.

A line-up of the nucleotide sequence of all the WT1 orthologs is also included (adapted from Clustal V output)
Human WT1 cDNA (Gessler et al., 1990)

```
TCAAGGCAGCCACACCCGGGGCTCTCCGCAACCCGACCGCCTGTCCGCTCCCCCAC
1   ---------------------------------------------------------- 60
AGTTCCTCGCCGCTGTTGCCCCCGAGGGCTGGCTGGGACGGGACGGGACGTA

SRQRPHTPGA LRRNPTACPPLPH

TTCCGGCCCTCTCCACCTTAACATTCAACCTACCCACCCACCCAGACGGGGAGG
61   ---------------------------------------------------------- 120
AAGGGCGGGGAGGGGAGGGTGGATGAGTAAGTGGGTGGGTGGGTCTCGGCCCTGCCG

CPPSLSPLPTTHPPRAGTA

GCCCCAGGGGCGGGCCGCACCTGTTCCACGACCTACCCAGGTGCCGAGGCCCAG
121   ---------------------------------------------------------- 180
CGGCTCCGGGGCCGCAGAGGGCCGCTAGGACCTGAAGGAGAACGATCG

AQAPGPRRLLAAILDQFLQ

AAGGGCGGGAGGGAGGGTGGATGAGTAAGTGGGTGGGTGGGTCTCGGCCCTGCCG
181   ---------------------------------------------------------- 240
CTGGGCCGAAGGTGCCACAGGGCCTCGGCCGCAGAGTCGTTGGCGAGGCCCGG

DPASTCVPPEPAQPHTLRSGP

CGGTCTACAGACGGTGGACTCGGAGTCGTTTACCCGAGGCTGACGCCCTGGGCC
241   ---------------------------------------------------------- 300
CTGGGCCGAAGGTGCCACAGGGCCTCGGCCGCAGAGTCGTTGGCGAGGCCCGG

GCLQQPQOQGVDRDPGGIGIWA

AATCCGCGGCGGGCTCCGGTCGCGACTTGCAGAGGTCCCGGCCTCCTCGCGCCCGCC
301   ---------------------------------------------------------- 360
CCCAGACTCGGAGTCGTTTACCCGAGGCTGCACGCCCTGGACTTGCGCGACGACGG

LGAASEAERLQGRRSRGAS

Translation Start→

GGGTCTGAGCCTCAGCAAATGGGCTCCGACGTGCGGGACCTGAACGCGCTGCTGCCCGCC
361   ---------------------------------------------------------- 420
CCCAGACTCGGAGTCGTTTACCCGAGGCTGCACGCCCTGGACTTGCGCGACGACGG

GSEPQQMGSDVDRDLNALPLA

GTCCCCCTCCTTGGGCTCCGGCGCGCATCTGATCCTGGCAGGCCCGGCGCGGACG
421   ---------------------------------------------------------- 480
CAGGGGAGGGACCAACCCGGGCCCAGCAAGGCAACGGCGACACTGGCCGCGGCTGAC

VPSLGGGGGCALPVSGAAQW

GGCGCCGGTGACTTGGCAGCTGGCCCGCCGCTGGCAGGTGGGCTGGGCTGGCCGGCC
481   ---------------------------------------------------------- 540
GCCGGCAGACCTGAAGCCGGGGCCGCCGGAAGCGCAGATGCCCAGCAACCCGGGCGG

APVLDFAPPAGASAYGSSLGPG

GGCGCCGGCCACGGCTCCGCACCCCCCCCCCCACCCGGCGCGGCGTACCTCCTTCAATGCCAAC
541   ---------------------------------------------------------- 600
GGCGCCGGTGCCGGCAGGGCGGCTGGGGGCGCGCGGCGCAGTGAGAAGTATGTTTGGT

APPPAPPPPPPPPHSFIKQ
```
Pig W71 cDNA

ACTTGGGAGCCACATTAAAAGGAGTGCTGCTGGAACGCTCCAGCTCAATGAAATGGACAG
-------------+----------------------------------------78
TGAATCCCTGCTGTAATTTTCTCAAGCGGACTCTTTGAGGCAGTGTGCAGTGGAGAAGCT
-------------+----------------------------------------138
19

L G A T K L G V A A G S S S S S M K W T E -
-------------+----------------------------------------
AAGGCGAGAAGCAACACCGAGGGTTGACAGAGCAGATACCCACGCCACCCGCAATCTCTC
-------------+----------------------------------------138

G Q S N H G T D E S T H A T P I L C -
-------------+----------------------------------------
GTGGTGGCCACGATCACAAATACACACACCACACGGTGCTTGTCAGGGGACTCAGGAGTCAG
-------------+----------------------------------------198

CACCACGGGCTATGCTTATGCTGTGCTCAGTAGAAGACAGACTCTTCTGACACGCT
-------------+----------------------------------------258

V P G V A P T L V R S A S E T S K R P -
-------------+----------------------------------------
CTCTCATGTTGGCTTTACGGCCAGCTTTGAGGCTGAGACGAGTCCTGAACAGAGGAA
-------------+----------------------------------------318

F M C A Y P G C N K R Y P K L S H L Q M -
-------------+----------------------------------------
TGATACCCCGAGGACACACTGTGGAGAAGCAGAAGCAGACTGGACGGAGAAGCGAGA
-------------+----------------------------------------378

H S R K H T G E K P Y Q C D F K D C E R -
-------------+----------------------------------------
GAAGGTTTCTCCTCAGACCAAGTCTCTCCAAGAGACACACCAAGAGACACACAGTGGA
-------------+----------------------------------------438

CTTCAAAGAAGAAGCTGGCTGAGTTTCTGATTTCTCTCTTGGCTGTGTGCTCAGACTTT
-------------+----------------------------------------
R F S R S D Q L K R H Q R R H T G V K P -
-------------+----------------------------------------
CATCCAGTGTAAACTGTGAGCGAAGATCTCCGCGAACACTGAGAACACACCACA
-------------+----------------------------------------498

G T A A G G T C A G A C T T T T C A A G A T G C G T T G G A C T C C T C T C G G
-------------+----------------------------------------
F Q C K T C Q R K F S R S D H L K T H T -
-------------+----------------------------------------
CCAGGACTCATACA
-------------+----------------------------------------512

GGTCCCTGAGTATGT
-------------+----------------------------------------
Marsupial mouse WTI cDNA, clone Sc41

AGCGTGCCTCCCCCGGCCTATGATGTCACACGCCCACCGACAGCTGCACCGGCAGCCAG

81

TCGACGGAGGGGCCAGATACCCTACAGTGTGCGGGTGCTGCAGCTGCGCCGTGC

S V P P P V Y G C H T P T D S C T G S Q

GCCCTGCTGTCCCCCGGATACCTACAGTGTGCGGGTGCTGAGTATGATGATGACT

141

CGGGACGAGGACGGCTGAGGGATGTTGTCACTGTTAGACATGTTTACTGAGCGCTGAT

A L L L R T P Y N S D N L Y Q M T S Q L

GAATGCAATGACCTGGAAACCAGATGACACTGGGAAAAAGGCCCCACACTGAAAGGGCCACACACACACAGGA

201

CTTACGTACTGACCTTTGCTTCTACTTGGATTCGGGCCTGAGTATATCTTTATAGTGTGGA

Y E N D N H T T P I L C A Q Y R I H T

CATGCCCTCTTTPAGAGAATPACAGAATGTGCCGGCCAGGTGGCAGGCGGCTCCGGCTCCCTACTATCT

261

GTACCGCAGAAATCTCTCTTTATGTTTCTACAGCGGCCTACGGGCCCAATGGGATGATAG

H G V F R G I D V R R V P G V A P T I

CGGGACGAGGACGGCTGAGGGATGTTGTCACTGTTAGACATGTTTACTGAGCGCTGAT

321

ACACTCAGCTGGCTACCTTGTTACTTTGGATCCCCGGTGTGACTTCCCGGTGTGGTGCTCT

V R S A T E T N E K R P F M C A Y P G C

AAACGAGGATACCTTTAAGCTGTGCTCCACTTACAGATGCATAGCAGGAAGCATACTGAGTGAAG

381

TTTTTGCTCTTCTAATGAAATCCGGACAGGGTGAATGTCTACGTATCGTCCTTCGTATGACACTC

N K R Y P K L S H L Q M H S R K H T G E

AACCCCTACCAGTTGACCTTTCAAGGACTGTGAAAGCCAGTTCTTCTGATCAGACCAACTTC

441

TTTGGGATGGTCACACTGTAAGTTCCTGACACTTGCCGCTAAGAGCTAGTCTGGTGGAG

K P Y Q C D F K D C E R R F S R S D Q L

AAACGACACCAAAAGGAGACACACAGGTTGGAACCAATTTCCAGTGAACCACTGTCAGAGA

501

TTTTGTGTGTTCTTCTGTGCTCCACACTTTTGTGTAAGGTCACTACATTTTGGAACAGTCTCCT

K R H Q R R H T G V K P F Q C K T C Q R

AAAGTTCTCCCGGTCTCGACACCTGAAGACACACACACAGGACTATACACAGGTAACCAAGT

561

TTCAAGAGGGCCAGACTGGTGAGCTTCTCCTGTGTCCTGTACGTATGTAACATTCTTGTCA

K F S R S D H L K T H T R T H T G K T S

236
Chick W71 cDNA, clone C2.1

CGGACAGGAGGGCCTGAGGCGGCTGCTGACACCCTTCCTATCCCCCCCGCGAC
1 -------------------------------------------------------- 60
RGHEGAEEAVGGPPPPYPPPPPRRH-

GCCGGGACAGACACACAGGCGGACACAGAGCAAGGCCAGCCAGAGACAGGCC
61 -------------------------------------------------------- 120
CGGCCCTGCTGTGCTGGCCGCTGGTGCTGGCTTGGCGCTGGCGCTGGCGCTGGG

ARRGAEAVpppp-

GCGCGGACAGACACACGGCCGCACACACGGACACGCGGCCACGGACACGCGCGAC
61 -------------------------------------------------------- 120
CGCGCCTGTCTGTGTGCCGGCGTGTGTGCCTGTGCGCCGGTGCCTGTGCGCGTGGGGT

ARTDTRPHRTTRGRHARTP-

CGCTGCTGCGCCGGCCGGGACAGCCCCGGAGGCGATCCTGGAGAACAACCTTTCCCTCGT
121 -------------------------------------------------------- 180
GGGACGCGACGCTGCCCCGTCGCGGCGCTCCGCTAGGACCTCTTGTTGGAAAGGGACAC

PCASAISHSPGDPGEQFPV-

CGCTGCTGCGCCGGCTCCCGCGGGCTGGACCCGGGGAGCTGCGTCGCGGAGCAAACCTCTCC
181 ---------+---------+---------+---------+---------+---------+ 240
GCGACAGGCGGCGGAGGGCGCCCAGCTGGCCGACGCGCCGGCGGCGCTGCTCCGTC

PCAGGAGTGCAGGACCGCCAGCGCGCCCATGCCGGTGAGCAGCGCGGCGCAGTGGGCTCCCGTC
241 ---------+---------+---------+---------+---------+---------+ 300
CGCTGCTGCGCCGGCCGGGACAGCCCCGGAGGCGATCCTGGAGAACAACCTTTCCCTCGT

KGAALRLVPAAGAGSGWRQRQ-

CTCCGACTCGAGAGGCCGACCACGGACGTCGGCCTCGCAGGCCGACCGCCGTCTCCGT
301 ---------+---------+---------+---------+---------+---------+ 360
GCGACCGAGGGGCGGCTGGTGCCTGCAGCCGGAGCGTCCGGCTGGCGGCAGAGGCAG

VSWKCGLGDDWLPGRQRKV-

Translation Start→

TGCGAGCAGATGGGGTCCGACGTCCGGGACCTGAACGCGCTGCTGCCCTCCGTGCCCTCC
361 ---------+---------+---------+---------+---------+---------+ 420
ACGCTCGTCTACCCAGGGTCGAGGGACTTGCCGCGCTGGCGCGAGCGCGGAGTCGCGC

CEQMSDVRLNALLLPSPVS-

CTACGCCGGCAGACACACTGCCCGCATGCCCGTGAGACCGGCCGGCCGATGGCTGGCGCCTCCGT
421 ---------+---------+---------+---------+---------+---------+ 480
GATGGCCGGTTGTCTGATCGACGGGCGTTACCGCCACTCGTGGCCGGCGTCCACCGAGGGACG

LPGNNSCAMPVSSAAQQWAPV-

CTGGAAGTTCCCGGGGGCCCTCTACGCGTCTGCGGCGGACCTCTCTCGATCGAGGAG
481 ---------+---------+---------+---------+---------+---------+ 540
GACCTGAGAAGGGGCGGCGGGAGGATGCCGAGCGAGCGCCGGGCTGTGAGGAATGAGTTCGT

LDSPPGASYGSLGPHSFIKQ-

GAACCCAGCTGGAACCGGCTTCGGACCCGCAGCAGCAGTACCTGAGCGCTTCCAGCGT
541 ---------+---------+---------+---------+---------+---------+ 600
CTTGGGTGACCTTTGCAGCCCTGGCGTGCTCCTCGTGACATGGAACCTGCGGAAGTGCAG

EPSWNGDSPHEEQYLSAFTV-
Alligator WTI cDNA, clone AL1

CAGGTTTACAGCAGTTGCCGTGGTTGACCGGCCCCCGAGCTACGGCCACGCTCCCTCGCAC
8

-----------

GTCCCAATGTCGTCCACCCCGCAACTCGGCCTTGCTGCCGCTGAGGGGACGGGCCCAGCTAC
86

-----------

Q G Y S T V A F D G P P S Y G H A P S H -

CACCCCGCCCGAGTTCTCCAACCTCTTCAAGCAAGAGGCCACATCGCCCCACGCAACGAGACT
68

-----------

GTGCAGCCGCTGCAAGAGGTGAGGAAGTTCGTGCTCCTGGGGTAGCGGGTCGTCTGAG
Q Y ST VA FD G pp Y G H A P S H

GACAGCTGCACGGGCGCAGCCCCCTCTGCTCTCGGGCTGCCACATCGCCCCACGCAACGAGACT
128

-----------

AGGGATCTCCTGTGCTGATGCACCGCAGGGGCCGCGGCCCACGCTCGGGGTGC
S L G D Q Q Y S V P P V Y G C H T

TCCTAGGAGACCAGACATCCTCGGTGCTCCCGCCCGCTGCCTGACCGCAGCCCCACGCAACGAGACT
188

-----------

GATGTGGACTGCCGCTGACGCCCCCCACGCTCGGGCTGCCACATCGCCCCACGCAACGAGACT
248

-----------

ATGGTTTACTGGAGGCTGAACTTACGTACTGTACCTTAGTTTACTTGAACCCTAGG
Y Q M T S Q L E C M T W N Q M N L G S T

TCCCTAGGAGACCAGACATCCTCGGTGCTCCCGCCCGCTGCCTGACCGCAGCCCCACGCAACGAGACT
308

-----------

AATTTCTCCGCTATCGCTCTCTATTTTTCTCTTTTTACTTGGGTGAGG
L K G H A T G Y E N E N H T A P M L Y S

TCCTAGGAGACCAGACATCCTCGGTGCTCCCGCCCGCTGCCTGACCGCAGCCCCACGCAACGAGACT
368

-----------

ACACCCTGGGGTTATGTCTTATGTGGGTACCTCAAAAATCTCCTTATGTTCTAGGCC
b C G A Q Y R I H T H G V F R G I Q D V R

CGAGCTGCCAGGAGTACGCTCCCAACTATTGTCGCCGATCAGCGAGACAAATGGAAAAACGT
428

-----------

GGGGAGTACACACGTATGGGACCGACATTATTCGCTATGAAATTCAATAGGGTAAATGTC
P F M C A Y P G C N K R Y F K L S H L Q

GTCAGGCCTCTCAGCGTATCCGATTGATACACACGACTCTCTCTCTCTCTCTCTCT
R V P G V A P T I V R S A S E T N E K R

GGGAGTACACACGCTATGGGACCGACATTATTCGCTATGAAATTCAATAGGGTAAATGG
P F M C A Y P G C N K R Y F K L S H L Q

ATGCACAGCAGAAAGCACACTGGTGAAAAACCATACCAGCAGGACTTTAAGGACTGAG
M H S R K H T G E K P Y Q C D F K D C E
Zebrafish WTI genomic DNA, clone Zia (exons in bold)

CGGCACGAGATAAAAAGCTAATCTGAACATATAACTGTCACCTACAATCAAACAGCTTGTGCT
1
GCCGTGCTCTATTGTGCAGATTGACTTTGTGTTATCTCTGGTTGGTCAAGGGAGACG

RHEITANLNITVTVYKSNACAG-
GTR*QLIT*LSPTNQLTLV-
ARDNS*SENHCLQIKRLC-

AGCCTGCAAGCTGTTGAGCCTGCTGTTGTATCTCTGGTTGGTCAAGGGAGACG

61

GCCGTGCTCTATTGTGCAGATTGACTTTGTGTTATCTCTGGTTGGTCAAGGGAGACG

SLQSCVAPAVSLSGSGSR-
ACRAVPSPLLCYPWVGQGRR-
PAELCSLACCVCILGWKDV-

TCTGCCCACTACAGTCGAGAAACAGGTGATTATTTGATTATAAGCATACACTACACTATC

121

TCGGACGTCTCGACACATCGGATCGGACGACACAATAGGAACCCACCCAGTTCCCTTCGTC

SATTVGTVYLVINST*HSF-
LPS*EQCILSIALNTHS-
CHYSRRNFSYQHLLTH-

ATTACTGTGCCAGCATACAGAAATGCGCATATGCTAAGCCTGATTTTTATATTATAGGACG

181

TAATGACACGGTGATGTCATATCAAATACATTGTATCTTGAGAATTCGACTCTGCTA

ITVPAYRNAIMLSASLRAD-
LCLHQTEMPLCALVNEQ-
YCASIQKCHYAKR*LIKSRW-

GGGCTTCAGCAGCGGGTAAATGAGGGAGACGGAACGGAACGAACTGAAATATA

241

CCGGAAGTGTCGCCCATTTACTCCTCCCTTGCTCCCTTGGCCCTCGTTGACCTTTATT

GLQRVRNVESGRNNGANVK-
GFSQSMREAGGTGERT*NK-
ASAGKGKREERGSEIRE-

GGTCTGCCCTAAGCTGCTGACATACGTTAGCATACGTATAGCAGCATCACGCTGAT

301

CCGACAGGATTGCAGCTGATTGGCAGTGTCAATCTCGAGTATAGTCGTCGTGCCACTA

GLPNVTVNRHS*SSYQQHGD-
VCLTS*LTVTVRAHISISSTVI-
SAR*RRD*PSQLESISAAR*S-

CGCATGATCATTTACTGCAAATGCTCTGGTACAGCATGCATTGATGTCTTGCTGGTACCA

361

GCGTACTAGTAAAGGTGTCATATGAGACTCTACATGCTGAGTATTGCTGCTGGTACCTA

RMIHYSLSMTMHAMMLFFF-
ASTFAULTV*PCML*CFFFF-
HDHSQLSQYDHCYDAFFFF-

TTTTGCAGATGTCTGGTTGCTGGTCAAGGATTTGCAATTGAATGACACTTTGT

421

AAAACGTCTACAGAAACAGCGACAGCAGCGTCAAATACCTGCTAAACAGCATCGTAACATTAGCCTGTAATCCTAATCTTGAGAACA
TGATTTCCTGCGGAGGTGCGGAGGATCCACACTCTGCGGATTTGCGGTTCAA

ACTAAAAGAGCCCTACACACTCCCTGACGTTAGTGAGACGGATACGGTAAAGCAGTT

CCGACACCAACGAAAAAGGCCTTACAGTGAGCGGCTACCCCTGTGAGAAGACTTGAA

CTCAGACCAGCTAAGACCCAGCATCAGTGAGGAGGATGCAAGGATCTCCATTTTTC

GCTATTGTATGCGATAACATACAGTATTCGTTTTACGATTCCAACAGCAATTGTACAAACCAACACACAGGTTAGAGATCTGGCAGGTTCTCCAGAACAT

AGTCTGGTCGATTTTGCTGTGGTCTCTTCTGTGTGTCCAAACGTGGTAACAGGTAAAAAC

ATAAAAGACAGCTACAGTACCTTGACTGCGCATTTGCTGTCACACTCTGCGGACTTCAATTTTCTTGTACTTCTTTATAACTGCTTGCTAAAGGACAACAAAATCAGTATAGAG

ATCAAACTGTGGGAGGAAGAACCATGCTAGTGGCAGTAAGCTGTTGATGATGAGAGAAGACACACAGGTTTGCACCATTGTCCATTTTGGGATAGTCACACTGAGTGTCTGACACCAGCGTCCAAGAGGTCT

S D Q L K R H Q R R H T G L H H C P F L -

Y F L V L L Y N C L L K D N K I S I E -

IFLYFFTITAC*RTTKSVA*FSCSTSL*LAKGQONQYR?
Unsequenced region ~2kb

AAAAGGTGTGATAGCGGTCTGCACTGTTGACTTGTCAACAAAGTTCACTTTCACTCGCGT
1 ------------------------------------------------------------- 60
TTTCCACACTACTGGCAGAGCTGACACTACATGTAAGTATGGTGAAGTGAGCGCA

K R C D S G L H C * L V N K V H F H S R -
K G V I A V C T V D L S T K F T F T P R V -
K V * R S A L T C Q S S L S L A F -

TCCAAGTTTCTGAACTCTCAATCAAAACACATTTAGCGTAGTGTTTGCTGATGTAATGCT
61 ------------------------------------------------------------- 120
AGGTTCAAAGACTTAGGAGTAGTTGTTGTTAATCGACACTCACAACAGACTCATAACCGA

S K F L N P Q S N T F S V S V C * C M A -
P S F * I L N Q T H L A * V F A D V W L -
Q V S E S I K H I * R E C L L M Y G * -

GAACACGCAGACTCTTGGTCTACATACATATAATTTTTTTTTTTTTTTTTTTTTTTTGTGCTGA
121 ------------------------------------------------------------- 180
CTTGTCGCCGCTGAGAGACGAGACTGTAATGATATTAAAAAAAAAAAAAAAAAAAAACACGACT

E Q A D S L G L T Y * F F F F F F F F L C * -
N R Q T L L V S H T N F F F F F F F C A D -
T G R L S W L I L I F F F F F V L T -

CACATTTCTGCTAGCGTTTGTGCTACTCTGTTCCCTGTTCTGTGTGTTGTGTTGTGTTGCTG
181 ------------------------------------------------------------- 240
GTGTAACACGATCCGCACAAAGACTGGAACAGACAGAAGAAGAAGAAGAACACACACACAC

H I W L G V S L D L C V L F L L C A C V -
T F G * A F L T C A S F C S C C V L V C -
H L A R R F * P V P R S V L V V C L C V -

TGTGCCGACACACTCTTGGTCTGTTCCCTGTTGTGTTGTGTTGAGCCTTAAGGCATCCACG
241 ------------------------------------------------------------- 300
ACACGCTTTGTTGTAACACAAACAGAGCAACACACACACACACACACACACACATCGCAATTCGCTAGT

C A T T L V C S C C C V C V G V K P F Q -
V R P H L C V P V V V C V * A L S H S S -
C D H T C V F L L L C V C V C R R * A I P V -

PCR sequence →

TGTGAACCTTGCTGACAGAAGAGGTTTTCACACTTGACACACACCTTTAAGACACCCCCGAC
301 ------------------------------------------------------------- 360
ACACCTTTGGAACAGACTCTCTTCCAAAGACTGGAACAGTCGTTGGAATCTCGGGTTCGCTGT

C E T C Q R K F S R S D H L K T H T R T -
V K P V R E S F H V Q T L R P T P G L -
* N L S E K V F T F T F P P * D P H Q D S -

CATACAGTTAACAACAGTTGCCTAAACCTTTTTCATTTTTTCATGATTGCTCTCTCCTTTT
361 ------------------------------------------------------------- 420
GTAGGCTCATTGTGTCAGCGGAAAGAATGAAAAGATACTAACAGCAGAGAGAAAA

H T G K T S A * T F S F F S * F A P L F -
I Q V K Q V K P F H F H F H D S L L S F -
Y R * N K C V N L F I F M R S S L S -

CCACCTTTTGATTTAAGAGTGTAATTTTTCAGGCTACTCTTTTTTCATCAGCTTGACTGG
421 ------------------------------------------------------------- 480
GGTGAGAAACTAAATTTCAACTTAAAAATGTCCAGTGAAAAAATAGAGTCGAACATGACC

PLF DLK LN F Y S I TF G H F I S A C T G -
TLF KV E FL Q V TF LSQ L V LG -

GAAGTTATAAAAAATAATTGAGCTGTTTTTTTTATATTGATAGCACCAGCTGTTTTTAGTATT

EVIKKIFAVVFLLILATGFSI -

KLKLYLQWFFY* *PLVLFVF

SY* KNC S G FFIDSHWF*YS -

CTCATTTCAACTGTAA

LIQL* -
SFNC?- -

HSTV? -

Zebrafish WTI predicted cDNA sequence

GATGTGGAGGAGTCGTCCATCCACACTCTCCTGTCGTCACCAGGACCAAGGAA

CTACACGCCTCTCATGCGCCCTAGTGAGGACGGTAACACGCAAGTTGGCTCTGGTTGCTT

D VR R V P G I T PA I VR S T E T N E -

AAAAGGCCTATTCTGCTGCTCCCTGCTGAGAAGAATTTATACTAGTGCAC

TTTTCCGGTAAGTACACCGGATGGGACCGACGTTGTTCTATAAAATTTGACAGCGTG

KRPFMCAYPGCNKRYFKLSH-

TTACAGATGCACACCCGTTAACACACAGGGAGAACCCCTATCAGTGTCATTCACAGAC

AATGCTCTACGGTGCCGATTTTGCTGTGGTCTCTTCTGTGTCTCG

LRQHSHSRKHTGEKPYQCDFTD-

TGTCGTCAGGTTTCCTCAGACACCCATCCACCAAGGAGAGAGAAGACACACACAGG

ACACCAAGTCCTCAGAGGATCTAGTCTGTCAGATTTCTGCTGTGTCATTCACACAGTGCAC

CGRFRSRSDDLKRHRQRRHTG-

PCR sequence

GTTAAGCCATTTCCCAGTCCAGAGAAACCTCTGCAGAAGGATTTTCAGTTCGCACCACCTTAC

CAATCGTGAAGTGCACACTTTGGAGCTCAGTCTCTTCTCACAAGAGTGAGGACAGTGACAG

VKPFQCETCQRCRFSRSDDLK-

ACCACCAACCCGGACACACATTACAGGTAACAA

TGGGTGTGGCCCTGTATGTCCATTTGTGT-
Alignment of the nucleotide sequence of the coding region of the WTI orthologs

- indicates no sequence change
. indicates a deletion/insertion
* indicates a conserved position in all species sequenced

Exon 1

<table>
<thead>
<tr>
<th>Species</th>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>ATGGGCTCGGACGTGCGGACCTGAACGCCTGCTGCCCGGCGTCCCTCCCTCGGCGTGGC</td>
</tr>
<tr>
<td>Mouse</td>
<td>-T-G-G-T-G-C-C-</td>
</tr>
<tr>
<td>Rat</td>
<td>-T-G-G-C-C-</td>
</tr>
<tr>
<td>Chick</td>
<td>-G-G-G-C-C-</td>
</tr>
</tbody>
</table>

Glycine 5 encoding region

<table>
<thead>
<tr>
<th>Species</th>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>G...CGGCAGCTTGCCCTGTGAGCGCGCGGGCGGCAAGGGCTGCTGCCCGGCGTGGCAC</td>
</tr>
<tr>
<td>Mouse</td>
<td>-CGG-----C-G-C-A-C-</td>
</tr>
<tr>
<td>Chick</td>
<td>-CAA-A-AG---A-G-A-T-C-</td>
</tr>
</tbody>
</table>

Proline 13 encoding region

<table>
<thead>
<tr>
<th>Species</th>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>TTTTGCCGCCGCTCGGCTTGGCTGGCTGGCTGGCGGCGGGCGGCACGCCGCCGACGCTGGGC</td>
</tr>
<tr>
<td>Mouse</td>
<td>-C-G-T--C-G-C-G-C-A-</td>
</tr>
<tr>
<td>Rat</td>
<td>-C-G-T--C-G-C-G-C-A-</td>
</tr>
<tr>
<td>Chick</td>
<td>-C-G-T--C-G-C-G-C-A-</td>
</tr>
</tbody>
</table>

Human

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTTTGCCGCCGCTCGGCTTGGCTGGCTGGCTGGCGGCGGCGGCACGCCGCCGACGCTGGAC</td>
</tr>
<tr>
<td>-CGG-----C-G-C-A-C-</td>
</tr>
<tr>
<td>-CAA-A-AG---A-G-A-T-C-</td>
</tr>
</tbody>
</table>

Mouse

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTTTGCCGCCGCTCGGCTTGGCTGGCTGGCTGGCGGCGGCGGCACGCCGCCGACGCTGGAC</td>
</tr>
<tr>
<td>-CGG-----C-G-C-A-C-</td>
</tr>
<tr>
<td>-CAA-A-AG---A-G-A-T-C-</td>
</tr>
</tbody>
</table>

Rat

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTTTGCCGCCGCTCGGCTTGGCTGGCTGGCTGGCGGCGGCGGCACGCCGCCGACGCTGGAC</td>
</tr>
<tr>
<td>-CGG-----C-G-C-A-C-</td>
</tr>
<tr>
<td>-CAA-A-AG---A-G-A-T-C-</td>
</tr>
</tbody>
</table>

Chick

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTTTGCCGCCGCTCGGCTTGGCTGGCTGGCTGGCGGCGGCGGCACGCCGCCGACGCTGGAC</td>
</tr>
<tr>
<td>-CGG-----C-G-C-A-C-</td>
</tr>
<tr>
<td>-CAA-A-AG---A-G-A-T-C-</td>
</tr>
</tbody>
</table>

Human

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>G...CGGCAGCTTGCCCTGTGAGCGCGCGGGCGGCAAGGGCTGCTGCCCGGCGTGGC</td>
</tr>
<tr>
<td>-T-G-G-T-G-C-C-</td>
</tr>
<tr>
<td>-T-G-G-C-C-</td>
</tr>
<tr>
<td>-G-G-G-C-C-</td>
</tr>
</tbody>
</table>

Mouse

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>G...CGGCAGCTTGCCCTGTGAGCGCGCGGGCGGCAAGGGCTGCTGCCCGGCGTGGC</td>
</tr>
<tr>
<td>-T-G-G-T-G-C-C-</td>
</tr>
<tr>
<td>-T-G-G-C-C-</td>
</tr>
<tr>
<td>-G-G-G-C-C-</td>
</tr>
</tbody>
</table>

Rat

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>G...CGGCAGCTTGCCCTGTGAGCGCGCGGGCGGCAAGGGCTGCTGCCCGGCGTGGC</td>
</tr>
<tr>
<td>-T-G-G-T-G-C-C-</td>
</tr>
<tr>
<td>-T-G-G-C-C-</td>
</tr>
<tr>
<td>-G-G-G-C-C-</td>
</tr>
</tbody>
</table>

Chick

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>G...CGGCAGCTTGCCCTGTGAGCGCGCGGGCGGCAAGGGCTGCTGCCCGGCGTGGC</td>
</tr>
<tr>
<td>-T-G-G-T-G-C-C-</td>
</tr>
<tr>
<td>-T-G-G-C-C-</td>
</tr>
<tr>
<td>-G-G-G-C-C-</td>
</tr>
</tbody>
</table>

Human

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCGTCATCCTCGGACAGCGGAGCGGAGCTGCTGCCCTGGCGCTGTCCCTCGGCCCTCCCGAGCAGCAG</td>
</tr>
<tr>
<td>-C-A-A---A---C---G---C---</td>
</tr>
<tr>
<td>-C-A-A---A---C---G---C---</td>
</tr>
<tr>
<td>-G-G-G---G---G---C---G---</td>
</tr>
</tbody>
</table>

Mouse

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCGTCATCCTCGGACAGCGGAGCGGAGCTGCTGCCCTGGCGCTGTCCCTCGGCCCTCCCGAGCAGCAG</td>
</tr>
<tr>
<td>-C-A-A---A---C---G---C---</td>
</tr>
<tr>
<td>-C-A-A---A---C---G---C---</td>
</tr>
<tr>
<td>-G-G-G---G---G---C---G---</td>
</tr>
</tbody>
</table>

Rat

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCGTCATCCTCGGACAGCGGAGCGGAGCTGCTGCCCTGGCGCTGTCCCTCGGCCCTCCCGAGCAGCAG</td>
</tr>
<tr>
<td>-C-A-A---A---C---G---C---</td>
</tr>
<tr>
<td>-C-A-A---A---C---G---C---</td>
</tr>
<tr>
<td>-G-G-G---G---G---C---G---</td>
</tr>
</tbody>
</table>

Chick

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCGTCATCCTCGGACAGCGGAGCGGAGCTGCTGCCCTGGCGCTGTCCCTCGGCCCTCCCGAGCAGCAG</td>
</tr>
<tr>
<td>-C-A-A---A---C---G---C---</td>
</tr>
<tr>
<td>-C-A-A---A---C---G---C---</td>
</tr>
<tr>
<td>-G-G-G---G---G---C---G---</td>
</tr>
</tbody>
</table>

Human

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGCCAGCCGGCTATTGCCAAATCAGG</td>
</tr>
<tr>
<td>-T-A---C---C---C---A---</td>
</tr>
<tr>
<td>-T-A---C---C---C---A---</td>
</tr>
<tr>
<td>-A-A---C---C---</td>
</tr>
</tbody>
</table>

Mouse

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGCCAGCCGGCTATTGCCAAATCAGG</td>
</tr>
<tr>
<td>-T-A---C---C---C---A---</td>
</tr>
<tr>
<td>-T-A---C---C---C---A---</td>
</tr>
<tr>
<td>-A-A---C---C---</td>
</tr>
</tbody>
</table>

Rat

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGCCAGCCGGCTATTGCCAAATCAGG</td>
</tr>
<tr>
<td>-T-A---C---C---C---A---</td>
</tr>
<tr>
<td>-T-A---C---C---C---A---</td>
</tr>
<tr>
<td>-A-A---C---C---</td>
</tr>
</tbody>
</table>

Chick

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGCCAGCCGGCTATTGCCAAATCAGG</td>
</tr>
<tr>
<td>-T-A---C---C---C---A---</td>
</tr>
<tr>
<td>-T-A---C---C---C---A---</td>
</tr>
<tr>
<td>-A-A---C---C---</td>
</tr>
</tbody>
</table>

249
<table>
<thead>
<tr>
<th>Exon 2</th>
<th>Human</th>
<th>GTTACAGCACGTGTCACCTTGGAGAGGAGGCCAGCTTACGGTCACACGGCCTCGCACCATTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>-A----T-G--------G-G-A---T---C---------------------T---C-</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>-A-----T---G-A---T---C-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>Chick</td>
<td>-G---G---C---G---T-----------T-G---------------C---G-T---C-</td>
<td></td>
</tr>
<tr>
<td>Alligator</td>
<td>*  ***  ****  *  **  *  **  **  ***  **  *  **  **  ***  *</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Human</th>
<th>CGGCCGAGTCTCCACACCCTCTGAGCAGGATAGGCTCCATGCCCATGGCCCAGGCTTGCGTGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>-----------------------------------T---C---A---C---G-----------------T---T---T</td>
</tr>
<tr>
<td>Rat</td>
<td>-A--------T---A---G-----------------T--G-----------------T--------T</td>
</tr>
<tr>
<td>Chick</td>
<td>-C--------T---A-------------------C---G---A---T---------------G---C-</td>
</tr>
<tr>
<td>Alligator</td>
<td>*  **********  *  **  **  ***  **  **  *  **  **  **  *  **  ****</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exon 3</th>
<th>Human</th>
<th>GTGACAGACGTACTCGGTGCCGCCCCCGGTCTATGGCTGCCACACCCCCACCGACAGCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>-C-----A--T-----G-----------------T--T-----T-----------------</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>-C-----A--T--A--G-----------------T--------T-----------------</td>
<td></td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td>GC-----T---A---T---G--------------T-------T---T---T---T</td>
<td></td>
</tr>
<tr>
<td>Chick</td>
<td>-G-----C-----------------C--T-----G--C-----T-----------GCG</td>
<td></td>
</tr>
<tr>
<td>Alligator</td>
<td>*  **  **********  *  **  *  **  **  ***  **  **  *  **  ***</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exon 4</th>
<th>Human</th>
<th>CCACACCACACCCTACCACACCCATAACCACACAACCCCCATCCT .......CTCCCCACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>----G-T-T ------------ T--C -------- CC-C ----------------- TT</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>----C----------------T--C-----------C-----------------TT</td>
<td></td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td>-----C------A--T--A-AT-----------C--T--G-----------G--T</td>
<td></td>
</tr>
<tr>
<td>Chick</td>
<td>----T-CG ----- A- -T--A-AT- -C-------CTC-T-----CT-ATACAC- --T--G-</td>
<td></td>
</tr>
<tr>
<td>Alligator</td>
<td>TGCA ----- ATAATC -------- TCTTCTATATACT-----</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exon 5</th>
<th>Human</th>
<th>AGTTGCTGCTGGGAGCTCCAGCTACGTCGCAAATGGCAGAGGCGAGGCAAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>-A-G---------------A-------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>-A-G---------------A-------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>-A-G---------------A-------------------------------------------</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exon 6</th>
<th>Human</th>
<th>CCACAGCACAGGGATACGAGGCAGATAACCACACAAAGCCCTTCCCTCT. . . . CTGCGGAGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>----G-T-T---------T--G-----GG-C--------T---T---T------------</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>----G--------T---T---G-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td>----C-----A--T-----A-AT-------G--T---G-----------------G--T-</td>
<td></td>
</tr>
<tr>
<td>Chick</td>
<td>----T-CG-----A--T--A-AT-------G-----GT--G---ATACAG---------T--G-</td>
<td></td>
</tr>
<tr>
<td>Alligator</td>
<td>TGCA-----A--T--A-AT-------G-----TG--T---GT--ATATAG---------T--</td>
<td></td>
</tr>
</tbody>
</table>

| ***  *  **  *  **  **  ***  **  *  **  *  **  **  ***  * |
### Exon 9

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
<th>Mouse</th>
<th>Rat</th>
<th>Marsupial Mouse</th>
<th>Chick</th>
<th>Alligator</th>
<th>Xenopus</th>
<th>Zebrafish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>GGTGTGAAACCATTCAGTAAACTTTGTCAGCGAAAAATCTCCTCCCGGTCCGACCACCTG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chick</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alligator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xenopus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zebrafish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Human**

AAGACCCACACCCAGGACTCATACAGGTAAAACAA

**Mouse**

---------------------------

**Rat**

--------------------------

**Marsupial Mouse**

-----A

**Chick**

----------

**Alligator**

------------

**Xenopus**

-------------------

**Zebrafish**

***********

### Exon 10

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
<th>Mouse</th>
<th>Rat</th>
<th>Marsupial Mouse</th>
<th>Chick</th>
<th>Alligator</th>
<th>Xenopus</th>
<th>Zebrafish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>GTGAAAAAGCCCTTCAGTGTGCGTGCGAGTTGCAGAAAGTTGCGCGTCCATGATG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marsupial Mouse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chick</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alligator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xenopus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zebrafish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Human**

AATTAGTCGCCACATCACAGAATGCATGAGAACAATGACAAAACTCCAGCTGCGTTGAG

**Mouse**

----G

**Rat**

-------------------

**Marsupial Mouse**

-----G-T-C

**Alligator**

---------------

---KTS---

**Human**

AAGACCCACACCCAGGACTCATACAGGTAAAACAA

**Mouse**

---------------------------

**Rat**

--------------------------

**Marsupial Mouse**

-----A

**Chick**

----------

**Alligator**

------------

**Xenopus**

-------------------

**Zebrafish**

***********

---KTS---
Appendix B
Structure prediction plots for WTI orthologs

Chou and Fasman prediction plots of α helix, β strand, and β turns were produced for WTI orthologs. Hydrophobicity profile using Kyte and Doolittle algorithms also were generated using pepplot (GCG). The plots for human and the most diverged sequences are shown for exon 1, exons 2-6 and exons 7-10.
Structure Prediction Profile for Human WT1, Exon 1

Chou & Fasman

Turn

Kyte-Doolittle
Structure Prediction Profile for Chick WT1, Exons 2-6
Structure Prediction Profile for Alligator WT1, Exons 2-6
Structure Prediction Profile for Human WT1, Exons 7-10

Chou & Fasman

Turn

Kyte-Doolittle
Structure Prediction Profile for Zebrafish WT1, Exons 7-9
Bibliography


