TOX3 : A CANDIDATE BREAST CANCER PREDISPOSITION GENE

Xenia Schmidt

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2012

Full metadata for this item is available in
St Andrews Research Repository
at:
http://research-repository.st-andrews.ac.uk/

Please use this identifier to cite or link to this item:
http://hdl.handle.net/10023/3104

This item is protected by original copyright
TOX3 – A Candidate Breast Cancer Predisposition Gene

Xenia Schmidt

This thesis is submitted in fulfilment for the degree of
Doctor of Philosophy at the
University of St. Andrews
School of Medicine

and

l'Université Bordeaux Segalen
Mention Sciences, Technologie, Santé – Option Biologie Cellulaire et Physiopathologie

November 2011
Abstract

Two thirds of breast cancers express the estrogen receptor alpha (ERα) and are estrogen-dependent for growth. In contrast, expression of ERα induces differentiation and senescence in normal human mammary epithelial cells. Both embryonic development and lineage commitment in the adult mammary gland are governed by transcriptional regulators, many of which have also been implicated in tumourigenesis. Genome-wide association studies have identified the previously uncharacterised putative transcription factor TOX high mobility group box family member 3 (TOX3) as a new candidate breast cancer susceptibility gene.

In the present study, I aimed to characterise TOX3 function in the normal human mammary epithelium and in breast cancer. Transcriptional profiling of human breast tumours showed that TOX3 was expressed in luminal but not in basal-like human breast tumours. Both in the normal human breast and mouse mammary gland, TOX3 expression appeared to be restricted almost exclusively to mature luminal epithelial cells, suggesting a function for TOX3 in luminal differentiation. In addition, TOX3 was shown to inhibit differentiation of bipotent MCF-10A cells along the myoepithelial lineage. Microarray analysis of MCF-7 luminal breast cancer cells expressing ectopic TOX3 showed that TOX3 expression resulted in the downregulation of luminal differentiation genes including PGR and GREB1. At the same time, TOX3 expression led to a marked induction of cell proliferation genes such as MYC and CCNA2. Using tandem affinity purification of TOX3 and subsequent mass spectrometric analysis, I identified the transcription/translation factor Y-box binding protein 1 (YBX1) as a potential TOX3 interacting protein. YBX1 has previously been shown to promote breast cancer progression through its function in a multitude of signalling pathways. The present work further showed that the nuclear protein TOX3 globally affected transcriptional regulation in an estrogen-independent manner. Finally, TOX3 promoted tumour growth in a luminal xenograft model.

Combined, my findings suggest that TOX3 regulates differentiation of normal human mammary epithelial cells along the luminal lineage by inhibiting myoepithelial differentiation. In tumour cells, TOX3 appears to suppress luminal differentiation and
promote proliferation, thereby allowing a switch from paracrine to autocrine signalling and transformation of ERα-positive cells.
Résumé

Deux tiers des cancers du sein expriment le récepteur à l’estrogène alpha (REα) et leur croissance dépend des estrogènes alors que l’expression de REα induit la différenciation et la sénescence des cellules humaines mammaires épithéliales normales. Le développement embryonnaire et la différenciation de la glande mammaire adulte sont contrôlés par des facteurs de transcription, dont beaucoup sont aussi impliqués dans la tumorigénèse. Plusieurs études d'association pan-génomiques ont identifié le putatif facteur de transcription TOX high mobility group box family member 3 (TOX3) comme un nouveau gène de prédisposition au cancer du sein.

L’objectif de cette étude était la caractérisation fonctionnelle de TOX3 dans l’épithélium mammaire normal et le cancer du sein. L’analyse des tumeurs du sein humaines par puces à ADN a montré que TOX3 est exprimé dans les tumeurs luminales mais pas dans les tumeurs basales. Dans le sein humain et la glande mammaire murine, l’expression de TOX3 est limitée exclusivement aux cellules épithéliales luminales matures, indiquant une fonction de TOX3 dans la différenciation luminale. De plus, j’ai démontré que TOX3 inhibe la différenciation myoépithéliale des cellules bipotentes MCF-10A. L’analyse microarray des cellules luminales cancéreuses MCF-7 exprimant TOX3 a montré que l’expression du facteur de transcription mène à une diminution d’expression des gènes de différenciation luminale comme PGR et GREB1 ainsi qu’à une augmentation de l’expression des gènes de prolifération comme MYC et CCNA2. Par la méthode de purification d’affinité en tandem (TAP) combinée à la spectrométrie de masse, j’ai identifié le facteur de transcription et traduction Y-box binding protein 1 (YBX1) comme potentiel partenaire protéique de TOX3. Or, il est connu que YBX1 favorise la progression tumorale du cancer du sein par son rôle dans une multitude des voies de signalisation.

De plus, les présents travaux ont montré que TOX3 est une protéine nucléaire qui influence globalement la régulation transcriptionnelle de manière estrogène-indépendante. Finalement, TOX3 accélère la formation des tumeurs dans un modèle de xénogreffe de cellules luminales.
En résumé, mes résultats proposent que TOX3 contrôle la différenciation luminaire des cellules humaines épithéliales normales en inhibant la différenciation myoépithéliale. Dans les cellules cancéreuses, TOX3 paraît de réprimer la différenciation luminaire et active la prolifération qui permet le passage d’une signalisation paracrine à une signalisation autocrine et la transformation des cellules REα-positives.
Declarations

1. Candidate’s declarations:

I, Xenia Schmidt, hereby certify that this thesis, which is approximately 55,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in July 2007 and as a candidate for the degree of Doctor of Philosophy in Medicine in July 2008; the higher study for which this is a record was carried out in the University of St Andrews and the University of Bordeaux between 2007 and 2011.

Date .................. Signature of Candidate ......................................

2. Supervisors’ declarations:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and the University of Bordeaux and that the candidate is qualified to submit this thesis in application for that degree.

Date .................. Signature of Supervisor, University of St Andrews ....................

Date .................. Signature of Supervisor, Université Bordeaux .........................

3. Permission for electronic publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as
required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisors regarding the electronic publication of this thesis:
Embargo on both [all] printed copy and electronic copy for the same fixed period of two years on the following ground: publication would preclude future publication.

Date .................. Signature of Candidate ...................
Signature of Supervisor, University of St Andrews ...............
Signature of Supervisor, Université Bordeaux ....................
Acknowledgements

First, I would like to thank Prof. Richard Iggo for giving me the opportunity to pursue my Ph.D. studies in his laboratory. I would like to thank him for sharing his vast knowledge, passion for cancer biology and close attention to detail.

I would like to thank Dr. Joanna Parish for being my co-supervisor in St. Andrews, for her insight, helpful suggestions and encouragement.

Thank you to Dr. Paul Reynolds for being a member of my postgraduate committee. Thank you as well to the members of my examination committee, Dr. Robert Clarke, Prof. Andrew Riches, Dr. Jean-Christophe Bourdon and Prof. Martin Teichmann.

I would like to thank all members of the Iggo lab, past and present, for their input, support and help. I would especially like to thank Sylvie André for welcoming me to the Iggo lab (and for teaching me how to do a proper Swiss raclette), Stephan Duss for sharing his expertise in mammary gland biology, Vita Fedele and CJ Osborne for much needed coffee breaks as well as Elodie Bachelard-Cascales, Elodie Richard, Laetitia Andrique and Stéphanie Verbeke for their help and input. Thank you to Dr. Gaëtan MacGrogan and Valérie Velasco at the Institut Bergonié for their assistance with immunohistochemistry and analysis of breast TMAs. I would further like to acknowledge the Plateforme Protéomique at the University of Bordeaux for mass spectrometric analysis and thank Benoît Rousseau and his colleagues at the animal facility at the University of Bordeaux for their help.

Thank you to the BMS gang in St. Andrews for making the lab a fun place to work. A big thank you to the fantastic people I have met at university, especially Verena, Helena, Gwen, Julia E. and Betty, for their support, insight and friendship, and for sharing my geeky passion for biology and cancer research.

I would like to thank Verena (again), Sabine, Meg, Gwen (again) and Anna and all my other friends for their support and encouragement over the years, and not least for being my life outside of science.

Je tiens à remercier Cécile pour son amitié et soutien, les filles de St Mich pour des soirées extraordinaires et Paul pour m’accueillir dans sa belle maison et pour partager sa passion pour le bon vin de Bordeaux. Je remercie Mélanie pour me faire découvrir et aimer Bordeaux, St Michel et la musique française.

Thank you to Clare for her love, encouragement and silliness and for being a source of strength.

Finally, I would like to thank my family, and my parents in particular, for their continued and unquestioning love and support. Mein Studium und die Doktorarbeit wären ohne euch nicht möglich gewesen. Danke.
Table of Contents

Abstract ............................................................................................................. i
Résumé ........................................................................................................... iii
Declarations ..................................................................................................... v
Acknowledgements .......................................................................................... vii
Table of Contents .............................................................................................. viii
List of Figures .................................................................................................... xii
List of Tables ..................................................................................................... xiv
List of Abbreviations ......................................................................................... xv

1 Introduction ..................................................................................................... 1

1.1 Breast cancer ............................................................................................... 1

1.2 Microarray-based classification of breast cancer ........................................ 1

1.3 Development and homeostasis of the mammary gland .............................. 3

1.3.1 Embryonic development of the mammary gland .................................. 4

1.3.2 Regulators of early mammary gland development and breast cancer .... 6

1.3.3 Structure and homeostasis of the adult mammary gland ....................... 7

1.3.4 The stem cell hierarchy in the mammary gland .................................... 8

1.3.5 Regulation of mammary gland development and homeostasis by steroid hormone receptors ................................................. 22

1.3.6 Regulators of stem cell activity, lineage commitment and differentiation ........................................................................................................... 25

1.4 Breast cancer subtypes and the cell of origin .............................................. 27

1.5 A model system for human breast cancer .................................................. 31

1.5.1 Human mammary epithelial cell culture ................................................. 32

1.5.2 Modelling human breast cancer by quantitative transformation of normal human mammary epithelial cells ................................................. 35

1.6 Identification of new mammary oncogenes and tumour suppressor genes .... 37

1.7 Thymocyte selection-associated high mobility group box protein family member 3 (TOX3) .............................................................................. 39

1.7.1 Identification of TOX3 as a breast cancer susceptibility gene in GWAS39

1.7.2 The high mobility group (HMG) box protein family ............................... 42

1.7.3 The thymocyte selection-associated HMG box (TOX) protein family ...45
2.4 RNA expression analysis .............................................. 72
   2.4.1 Isolation and reverse transcription of RNA .................. 72
   2.4.2 Quantitative PCR .............................................. 72
   2.4.3 Microarray analysis ............................................ 73
2.5 Protein expression analysis .......................................... 73
   2.5.1 TOX3 antibodies ............................................. 73
   2.5.2 Preparation of whole cell extracts and protein quantitation 73
   2.5.3 Western blot analysis ........................................ 74
   2.5.4 Co-immunoprecipitation ..................................... 74
   2.5.5 Immunofluorescent staining .................................. 75
   2.5.6 Immunohistochemistry ....................................... 76
2.6 Tandem affinity purification (TAP) of TOX3 ....................... 76
   2.6.1 TAP .......................................................... 76
   2.6.2 Optimisation of TAP conditions ............................. 78
   2.6.3 Silver stain analysis ......................................... 78
Results

3.1 Expression of TOX3 in the normal breast and breast cancer

3.1.1 TOX3 is expressed in the luminal cluster in ER-positive breast tumours

3.1.2 Immunohistochemical analysis of tissue microarrays confirms TOX3 expression in luminal tumours

3.1.3 A model system for endogenous TOX3 expression

3.1.4 TOX3 expression in the normal breast epithelium

3.2 Identification of TOX3 interacting proteins by tandem affinity purification

3.2.1 Lentiviral NTAP expression constructs

3.2.2 Optimisation of TAP purification conditions

3.2.3 Tandem affinity purification of TOX3

3.2.4 Mass spectrometric analysis of purified TOX3 and interacting proteins

3.2.5 Analysis of TOX3 protein interaction partners

3.2.6 Affinity purification of rabbit polyclonal TOX3 antibody

3.2.7 Knockdown of endogenous TOX3 using shRNA

3.3 Identification of TOX3 target genes and signalling pathways

3.3.1 Lentiviral expression vectors

3.3.2 Microarray analysis of TOX3 target genes

3.3.3 TOX3 represses a subset of ER target genes

3.3.4 TOX3 has a global effect on transcription

3.3.5 TOX3 localises to the nucleus

3.4 A role for TOX3 in lineage commitment of mammary epithelial cells

3.4.1 Culture of human mammary epithelial cells from organoids

3.4.2 Propagation of stem and progenitor cells in mammosphere culture

3.4.3 Propagation of luminal mammary epithelial cells in adherent culture

3.4.4 BPEC culture does not prevent the induction of p16

3.4.5 TOX3 expression decreases the myoepithelial subpopulation

3.4.6 Expression of TOX3 does not change ESR1 expression in BPEC

3.4.7 Three-dimensional culture of TOX3-expressing cells

3.5 A TOX3 tumour model

3.5.1 TOX3 expression confers short-term growth advantage but does not alter long-term proliferation of MCF-7 cells

3.5.2 TOX3 increases the tumour take in the MCF-7 luminal tumour model

3.5.3 Quantitative transformation of normal mammary epithelial cells
Discussion ........................................................................................................... 151

4.1 TOX3 is a luminal gene .............................................................................. 152

4.2 A transcriptional network that regulates luminal epithelial development and differentiation ............................................................................................................ 153

4.2.1 TOX3 is a nuclear protein ...................................................................... 153
4.2.2 Transcriptional regulators of luminal differentiation .............................. 155
4.2.3 TOX3 and ESR1 ................................................................................... 157
4.2.4 TOX3 and CITED1 ................................................................................ 158
4.2.5 TOX3 potentially interacts with YBX1 ................................................ 159

4.3 A role for TOX3 in luminal breast cancer .............................................. 162

4.3.1 TOX3 is expressed in luminal breast tumours ................................. 162
4.3.2 TOX3 induces proliferation and represses luminal differentiation genes in MCF-7 cells – a tentative interpretation ......................................................... 162
4.3.3 Creation of a luminal tumour model by quantitative transformation of normal human mammary epithelial cells ....................................................... 166
4.3.4 The predicted non-coding RNA gene LOC643714 .............................. 170

4.4 How does TOX3 cause breast cancer? .................................................... 170

4.5 Concluding remarks and future perspectives ........................................ 171

References ......................................................................................................... 173

Appendix ............................................................................................................. 197

6.1 Plasmid vectors used in this study .............................................................. 197

6.2 Antibodies used in this study .................................................................. 200

6.3 Ethics approvals ....................................................................................... 202

6.4 Mass spectrometry .................................................................................... 207

6.4.1 Mass spectrometry specifications and data analysis .......................... 207
6.4.2 Mass spectrometry results including keratins .................................... 210

6.5 Gene set enrichment analysis of the TOX3 microarray .......................... 213
List of Figures

Figure 1. Cross-section of a mammary duct................................................................. 9
Figure 2. Mouse and human stem and progenitor cell assays................................. 13
Figure 3. Epithelial cell hierarchy in the mammary gland........................................... 15
Figure 4. The cell of origin of the different breast cancer subtypes......................... 30
Figure 5. The TOX3 locus............................................................................................ 40
Figure 6. Model of HMG-box structure and function............................................... 44
Figure 7. Conservation of the HMG-box motif......................................................... 49
Figure 8. TOX3 is evolutionarily conserved across species...................................... 50
Figure 9. TOX protein domains.................................................................................. 52
Figure 10. TOX function in thymocyte lineage commitment.................................... 56
Figure 11. TOX3 is expressed in the luminal cluster in ER-positive tumours........... 84
Figure 12. Optimisation of TOX3 immunohistochemistry....................................... 85
Figure 13. TOX3 is expressed in luminal but not in basal tumours....................... 86
Figure 14. TOX3 expression in 51 breast cancer cell lines....................................... 87
Figure 15. Western blot analysis of TOX3 expression in breast cancer cell lines.... 88
Figure 16. TOX3 expression in the normal human and mouse mammary gland.... 91
Figure 17. Tandem affinity purification protocol....................................................... 94
Figure 18. NTAP-TOX3 lentiviral expression vectors.............................................. 96
Figure 19. Western blot analysis of NTAP constructs............................................... 97
Figure 20. The hUbiC lentiviral vector produces a TOX3-GFP fusion protein........ 98
Figure 21. The CMV-NTAP control lentiviral vector produces an NTAP-GFP fusion protein.............................................................................................................. 100
Figure 22. Tandem affinity purification pilot experiment....................................... 102
Figure 23. Optimisation of TAP lysis and elution conditions................................... 103
Figure 24. Silver stain and Western blot analysis of NTAP-TOX3 TAP................... 105
Figure 25. Transient expression of NTAP-TOX3 induces heat shock proteins..... 107
Figure 26. Immunoprecipitation of TOX3 ................................................................. 109
Figure 27. Epitopes recognized by TOX3 antibodies.............................................. 110
Figure 28. Custom-made TOX3 antibody................................................................. 111
Figure 29. Affinity-purification of polyclonal rabbit TOX3 Rb57 antibody........... 113
Figure 30. Knockdown of endogenous TOX3 using shRNAmir vectors............... 115
Figure 31. TOX3 lentiviral expression vectors......................................................... 117
Figure 32. TOX3 expression in transduced MCF-7 cells..............................118
Figure 33. TOX3 microarray. ......................................................................120
Figure 34. TOX3 target gene expression......................................................122
Figure 35. TOX3 acts as a global transcriptional regulator............................123
Figure 36. TOX3 localises to the nucleus......................................................126
Figure 37. In vitro stem cell and progenitor assays. ......................................128
Figure 38. Organoid culture of human mammary epithelial cells...................130
Figure 39. Propagation of stem and progenitor cells in mammosphere culture..131
Figure 40. Human mammary epithelial cells in adherent culture...................133
Figure 41. Growth properties of breast primary epithelial cells......................135
Figure 42. TOX3 expression inhibits myoepithelial differentiation..................138
Figure 43. TOX3 expression in breast primary epithelial cells (BPEC)..........139
Figure 44. TDLU assay. .............................................................................141
Figure 45. Stable expression of TOX3 in MCF-7 and MDA-MB-231 cell..........143
Figure 46. Intraductal TOX3 tumour model..............................................145
Figure 47. Protocol for the creation of a luminal tumour model......................147
Figure 48. Quantitatively transformed BPEC form tumours in NSG mice.......149
Figure 49. Model of TOX3 function in normal and tumour cells..................171
Figure 50. Maps of lentiviral vectors used for tandem affinity purification.......199
List of Tables

Table 1. Cell surface markers used for mammary epithelial cell sorting..................18
Table 2. Human TOX family members and their murine homologues.......................47
Table 3. PCR primers used for TAP-TOX3 and TAP control cloning .......................70
Table 4. PCR primers used for TOX3-GFP cloning ............................................70
Table 5. PCR primers used for HA-TOX3 cloning ..............................................71
Table 6. Primers used for quantitative PCR .......................................................72
Table 7. TOX3 status of breast cell lines ............................................................89
Table 8. Conserved genes in the mature luminal epithelial subpopulation ..............92
Table 9. Mass spectrometry analysis of proteins associated with NTAP-TOX3 ...106
Table 10. Tumour formation in NSG mice .........................................................148
Table 11. Plasmid vectors used in this study ......................................................197
Table 12. Antibodies used in this study ..............................................................200
Table 13. Mass spectrometry results including keratins ....................................210
Table 14. Top ranked genes in TOX3 arrays .......................................................213
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTA2</td>
<td>Actin, alpha 2, smooth muscle, aorta</td>
</tr>
<tr>
<td>AGR2</td>
<td>Anterior gradient homolog 2 (<em>Xenopus laevis</em>)</td>
</tr>
<tr>
<td>bFGF</td>
<td>Basic fibroblast growth factor</td>
</tr>
<tr>
<td>BMI1</td>
<td>BMI1 polycomb ring finger oncogene (B lymphoma Mo-MLV insertion region 1 homolog)</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>BPEC</td>
<td>Breast primary epithelial cell</td>
</tr>
<tr>
<td>BRCA1/2</td>
<td>Breast cancer susceptibility gene 1/2</td>
</tr>
<tr>
<td>CAGF9</td>
<td>CAG trinucleotide repeat-containing gene F9 protein</td>
</tr>
<tr>
<td>cAMP</td>
<td>3′-5′-cyclic adenosine monophosphate</td>
</tr>
<tr>
<td>CBB</td>
<td>Calmodulin binding buffer</td>
</tr>
<tr>
<td>CBP</td>
<td>Calmodulin binding peptide</td>
</tr>
<tr>
<td>CBP</td>
<td>CREB-binding protein</td>
</tr>
<tr>
<td>CE</td>
<td>Calmodulin eluate</td>
</tr>
<tr>
<td>CEB</td>
<td>Calmodulin elution buffer</td>
</tr>
<tr>
<td>CGH</td>
<td>Comparative genome hybridisation</td>
</tr>
<tr>
<td>CHEK2</td>
<td>CHK2 checkpoint homolog 2 (<em>Schizosaccharomyces pombe</em>)</td>
</tr>
<tr>
<td>CITED1</td>
<td>Cbp/p300-Interacting transactivator 1</td>
</tr>
<tr>
<td>CMV</td>
<td>Cytomegalovirus</td>
</tr>
<tr>
<td>CRE</td>
<td>cAMP response element</td>
</tr>
<tr>
<td>CREB</td>
<td>cAMP response element-binding protein</td>
</tr>
<tr>
<td>DAPI</td>
<td>4′,6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>DCIS</td>
<td>Ductal carcinoma <em>in situ</em></td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle Medium</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>E2</td>
<td>17β-Estradiol</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular matrix</td>
</tr>
<tr>
<td>EGF(R)</td>
<td>Epidermal growth factor (receptor)</td>
</tr>
<tr>
<td>EMT</td>
<td>Epithelial-mesenchymal transition</td>
</tr>
<tr>
<td>ERα, ESR1</td>
<td>Estrogen receptor alpha</td>
</tr>
<tr>
<td>ERE</td>
<td>Estrogen receptor response element</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>ERBB2</td>
<td>v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence activated cell sorting</td>
</tr>
<tr>
<td>FGFR2</td>
<td>Fibroblast growth factor receptor 2</td>
</tr>
<tr>
<td>FOXA1</td>
<td>Forkhead box A1</td>
</tr>
<tr>
<td>GATA3</td>
<td>GATA-binding protein 3</td>
</tr>
<tr>
<td>GREB1</td>
<td>Growth regulation by estrogen in breast cancer 1</td>
</tr>
<tr>
<td>GnRH</td>
<td>Gonadotropin-releasing hormone</td>
</tr>
<tr>
<td>GUS</td>
<td>β-Glucuronidase</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association study</td>
</tr>
<tr>
<td>IRS1</td>
<td>Insulin receptor substrate 1</td>
</tr>
<tr>
<td>H&amp;E</td>
<td>Haematoxylin and eosin</td>
</tr>
<tr>
<td>HBSS</td>
<td>Hank’s balanced salt solution</td>
</tr>
<tr>
<td>HEK 293T</td>
<td>Human embryonic kidney 293T cells</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid</td>
</tr>
<tr>
<td>HER2</td>
<td>Human epidermal growth factor receptor 2 (see ERBB2)</td>
</tr>
<tr>
<td>HMEC</td>
<td>Human mammary epithelial cell</td>
</tr>
<tr>
<td>HMF</td>
<td>Human mammary fibroblast</td>
</tr>
<tr>
<td>HMG</td>
<td>High mobility group</td>
</tr>
<tr>
<td>HMM</td>
<td>Human mammary cell medium</td>
</tr>
<tr>
<td>HUGO</td>
<td>Human Genome Organisation</td>
</tr>
<tr>
<td>IGF1</td>
<td>Insulin-like growth factor 1</td>
</tr>
<tr>
<td>KRT14, K14, CK14</td>
<td>Cytokeratin 14</td>
</tr>
<tr>
<td>KRT18, K18, CK18</td>
<td>Cytokeratin 18</td>
</tr>
<tr>
<td>kb</td>
<td>Kilo base pairs</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilo Dalton</td>
</tr>
<tr>
<td>LCK</td>
<td>Lymphocyte-specific protein tyrosine kinase</td>
</tr>
<tr>
<td>LEF1</td>
<td>Lymphoid enhancer-binding factor 1</td>
</tr>
<tr>
<td>LGR5</td>
<td>Leucine rich repeat containing G protein coupled receptor 5</td>
</tr>
<tr>
<td>LOH</td>
<td>Loss of heterozygosity</td>
</tr>
<tr>
<td>LSP1</td>
<td>Lymphocyte-specific protein 1</td>
</tr>
<tr>
<td>Ma-CFC</td>
<td>Mammary colony-forming cell</td>
</tr>
<tr>
<td>MAP3K1</td>
<td>Mitogen activated protein 3 kinase 1</td>
</tr>
<tr>
<td>MaSC</td>
<td>Mammary stem cell</td>
</tr>
</tbody>
</table>
List of Abbreviations

MHC | Major histocompatibility complex  
mRNA | Messenger RNA  
MRU | Mammary repopulating unit  
MMTV | Mouse mammary tumour virus  
MSPM | Mammosphere medium  
MUC1 | Mucin 1, cell surface associated  
MYC | v-myc myelocytomatosis viral oncogene homolog (avian)  
NCBI | National Center for Biotechnology Information  
NLS | Nuclear localisation signal  
NRG3 | Neuregulin 3  
PALB2 | Partner and localizer of BRCA2  
PBS | Phosphate buffered saline  
PGK | Phosphoglycerate kinase  
PGR, PR | Progesterone receptor  
PLK1 | Polo-like kinase 1  
PNUT | Phosphatase nuclear targeting subunit  
PRL(R) | Prolactin (receptor)  
PTEN | Phosphatase and tensin homolog  
PyMT | Polyoma Middle T antigen  
RAG2 | Recombination activating gene 2  
RANK(L) | Receptor activator of NF-kappaB (ligand)  
RNA | Ribonucleic acid  
RUNX3 | Runt related transcription factor 3  
SBB | Streptavidin binding buffer  
SBP | Streptavidin binding peptide  
SE | Streptavidin eluate  
SEB | Streptavidin elution buffer  
SDS-PAGE | Sodium dodecyl sulfate polyacrylamide gel electrophoresis  
SMA | α-smooth muscle actin (see ACTA2)  
SMARCE1 | SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1  
SNP | Single nucleotide polymorphism  
SP | Single positive  
SOX | SRY (sex determining region Y)-box
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRY</td>
<td>Sex determining region Y</td>
</tr>
<tr>
<td>TAP</td>
<td>Tandem affinity purification</td>
</tr>
<tr>
<td>TCA</td>
<td>Trichloroacetic acid (TCA)</td>
</tr>
<tr>
<td>TCR</td>
<td>T cell receptor</td>
</tr>
<tr>
<td>TEB</td>
<td>Terminal end bud</td>
</tr>
<tr>
<td>TERT</td>
<td>Telomerase reverse transcriptase</td>
</tr>
<tr>
<td>TDLU</td>
<td>Terminal ductal lobular unit</td>
</tr>
<tr>
<td>TMA</td>
<td>Tissue microarray</td>
</tr>
<tr>
<td>TNRC9</td>
<td>Trinucleotide containing 9</td>
</tr>
<tr>
<td>TOX</td>
<td>Thymocyte selection-associated high mobility group box</td>
</tr>
<tr>
<td>TOX3</td>
<td>TOX high mobility group box family member 3</td>
</tr>
<tr>
<td>TP53, p53</td>
<td>Tumour protein p53</td>
</tr>
<tr>
<td>UBF1</td>
<td>Upstream binding factor 1 (see UBTF)</td>
</tr>
<tr>
<td>UBTF</td>
<td>Upstream binding transcription factor, RNA polymerase I</td>
</tr>
<tr>
<td>WNT</td>
<td>Wingless-type MMTV integration site family</td>
</tr>
<tr>
<td>XBP1</td>
<td>X-box binding protein 1</td>
</tr>
<tr>
<td>YBX1, YB-1</td>
<td>Y-box binding protein 1</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Breast cancer

Breast cancer is the most common type of cancer in women in the industrialized world. It is estimated to have accounted for 28% of all female cancer incidences and 17% of all cancer deaths in women in Europe in 2008 (Ferlay et al., 2010).

Breast cancer is a very heterogeneous disease both at the molecular and histopathological levels, which is reflected in its clinical behaviour. Several distinct subtypes have been described based on gene expression profiling (discussed below). In addition, individual breast tumours consist of populations of cells with distinct phenotypes and molecular profiles, which is already evident at the stage of ductal carcinoma in situ (DCIS) before a tumour becomes invasive (Allred et al., 2008).

The current focus and prevailing notion in the field of breast cancer research is that the distinct breast cancer subtypes originate in different types of mammary epithelial cells that serve as the “cell of origin”. The identification of the cell of origin of the different tumour types is closely related to the question of the epithelial cell hierarchy in the normal mammary gland, and both require the identification and characterisation of the signalling proteins and transcription factors that govern mammary gland development by controlling progression along the mammary lineage. A better understanding of tumour heterogeneity is ultimately the basis for the development of effective and targeted treatments for the different tumour subtypes.

1.2 Microarray-based classification of breast cancer

In the last decade, gene expression profiling of breast tumours has led to the discovery of molecular subtypes of invasive ductal carcinoma (Farmer et al., 2005; Herschkowitz et al., 2007; Perou et al., 2000; Sorlie et al., 2001; Sotiriou et al., 2003). Expression of estrogen receptor alpha (ER/ESR1) is a major determinant of the tumour subtype, and tumours can broadly be classified into subtypes which express ER together with a set of genes that are characteristic of luminal epithelial cells of the mammary gland (luminal tumour subtypes), and those that are negative for the expression of these luminal genes (basal-like tumours and human epidermal growth factor receptor 2 (HER2/ERBB2)-positive tumours). The luminal class can be
subdivided further into the luminal A and luminal B subtypes. The luminal A subtype is characterised by high expression of ER and its target gene progesterone receptor (PGR/PR) and by the expression of a cluster of genes including forkhead box A1 (FOXA1), X-box binding protein 1 (XBP1) and GATA-binding protein 3 (GATA3), termed the luminal or ER cluster. Luminal B tumours generally exhibit low to moderate expression of ER and luminal specific genes, including the ER cluster, have higher proliferative rates and are of higher grade (Sorlie et al., 2001; Sorlie et al., 2003). In the clinic, the classification of luminal B tumours has been cause for controversy as ER+/ERBB2+ tumours are considered luminal B by some but not by others (Bhargava and Dabbs, 2008; Carey et al., 2006; Perou, 2010). The basal-like tumour subgroup is characterised by the expression of a set of genes that are characteristic of basal epithelial cells, including basal keratins, and lacks expression of ER, PR, the luminal gene cluster, as well as ERRB2 (Sorlie et al., 2003). In the clinical pathological setting, the basal subtype has traditionally been referred to as triple-negative breast cancer due to the lack of ER, PR and ERBB2 expression as defined by immunohistochemical staining, although the direct equivalence of these classifications is controversial (Gusterson, 2009). Sorlie and colleagues have described the ERBB2-positive tumour subtype as characterised by high expression of several genes in the ERBB2 amplicon at 17q22.24, yet some tumours classified within the ERBB2-positive subgroup lack ERBB2 expression, and vice versa, suggesting that ERBB2 expression does not accurately describe the non-basal ER-negative subgroup. This is consistent with findings from our group following the publication of the original classification by Sorlie et al. Our group identified a new subgroup of tumours, termed “molecular apocrine” (Farmer et al., 2005), which is characterised by expression of the androgen receptor and the lack of ER expression. This group at least partially overlaps with the HER2-positive subtype, thereby allowing for a more accurate classification of breast tumours. Thus, this alternative classification includes the luminal, molecular apocrine and basal subtypes. A recent follow-up study by Perou and colleagues identified the claudin-low subtype, which is characterised by low expression of genes involved in tight junctions and cell-cell adhesion, including claudins 3, 4, 7 and E-cadherin, as well as absence of ER, PR and ERBB2 expression. Claudin-low tumours further show low to absent expression of luminal genes, high levels of epithelial-to-mesenchymal transition (EMT) markers, immune response genes and inconsistent expression of basal genes (Herschkowitz et
al., 2007; Prat et al., 2010). Notably, the claudin-low profile resembles the signature of putative mammary stem cells (Hennessy et al., 2009; Prat et al., 2010). It appears that the remaining normal-like breast cancer subtype does not represent a clinically relevant subtype but stems from poorly sampled tumour tissue (Weigelt et al., 2010). It is of note that the classification of breast tumours by expression profiling is largely based on the tumour phenotype and the expression pattern of the (normal) cell type that it resembles rather than the genotype. The only exception to this is the group of HER2-positive tumours, which are classified based on an oncogenic mutation, or genotype. As mentioned above, our group has previously suggested the term “molecular apocrine” to describe this subtype of tumours, reflecting the phenotype of this breast cancer subtype which is characterised by increased androgen signalling in an ER-negative luminal cell, rather than its genetic make-up (Farmer et al., 2005).

It is not surprising that the heterogeneous nature of breast cancer has important clinical implications for prognosis and treatment of the disease. Several of the molecular profiling studies have reported that the definition of a tumour subtype by signature genes allows for the prediction of critical clinical parameters such as overall prognosis, treatment response, and therapeutic resistance (Sorlie et al., 2003; van't Veer et al., 2002; West et al., 2001).

1.3 Development and homeostasis of the mammary gland

Over the last decades important progress has been made in the understanding of signalling pathways that are involved in the morphogenesis of the breast, but the molecular mechanisms underlying the different developmental stages are still not fully understood. My discussion will be limited to the key events, structures and regulatory pathways involved in mammary gland development and structure in humans and mice. A more in-depth discussion of the current knowledge of breast development and branching morphogenesis can be found in a number of recent reviews (Hennighausen and Robinson, 2001; Hens and Wysolmerski, 2005; Howard and Gusterson, 2000; Lu et al., 2006; Mikkola and Millar, 2006; Sternlicht, 2006). Expression, knockout, and transplantation studies have provided insight into mammary gland development in mice, whereas less is known about human breast development owing to constraints on tissue availability and the lack of experimental
Introduction

systems. While key events in the early development of the murine and human mammary gland are believed to be similar, the morphology of the adult breast differs markedly between the species, as addressed below.

The mammary gland consists of an extensive tree-like system of branched ducts that is composed of the mammary epithelium, which in turn is embedded in the surrounding stroma. Unlike other branched organs, the mammary gland develops predominantly after birth in defined stages that are connected to sexual development and reproduction. The stages of mammary development are embryogenesis, prepubertal development, puberty, pregnancy, lactation, and involution. During embryogenesis, a rudimentary ductal tree develops. After birth, the mammary gland grows mainly isometrically with the rest of the body until puberty, when the ductal tree undergoes extensive branching morphogenesis in response to systemic hormonal cues. During adult life, the branched network of ducts is refined in response to cyclic ovarian stimulation. Development is governed by the intricate interplay of epithelial-mesenchymal interactions, systemic steroid and peptide hormone signalling pathways and transcriptional networks.

1.3.1 Embryonic development of the mammary gland

The two main compartments forming the mammary gland are the epithelium and the surrounding stroma, which are derived from the embryonic ectoderm and mesoderm, respectively.

Early embryonic development of the mammary gland resembles the formation of other skin appendages such as hair follicles and teeth, all of which are controlled by epithelial-mesenchymal interactions and are independent of systemic hormonal cues (Mikkola and Millar, 2006). However, the notion that the mammary gland is a modified apocrine or sebaceous gland is controversial (Howard and Gusterson, 2000). In the mouse, mammary development begins at mid-gestation with the formation of bilateral milk lines along the ventrolateral surface ectoderm that appear in response to mesenchymal signals on embryonic day (E) 10.5. At this stage, epidermal cells within the milk line become multilayered and columnar, thus protruding from the single-layered periderm. On E11.5, five pairs of placodes, local thickenings, form along the
mammary lines. This is thought to occur through migration of cells within the mammary line. Placode formation is followed by the invagination of cells into the underlying mesenchyme to form bulb-shaped mammary buds, the primary mammary rudiment or anlage, between E11.5 and E12.5. At the same time, mesenchymal cells surrounding the buds condense and differentiate to give rise to a dense mammary mesenchyme. Differentiation of the mesenchyme is characterised by expression of the androgen receptor, and in many mouse strains the mammary buds eventually degenerate in response to foetal androgen signalling in male embryos. In female embryos, the epithelial cells start proliferating on E15.5-E16.5 after a period of morphological quiescence, and the buds extend through the mesenchyme into the preadipocytes of the developing fat pad, forming a mammary sprout that starts to branch and forms a lumen in response to signalling from the stromal environment of the mammary fat pad. This gives rise to the rudimentary ductal tree consisting of a primary duct and 15-20 secondary ducts that are present at birth (Hens and Wysolmerski, 2005; Lu et al., 2006). Unlike branching morphogenesis in the adult mammary gland, embryonic branching is induced by epithelial-mesenchymal interactions and occurs in estrogen receptor (ER) knockout mice, demonstrating its independence from hormonal control by estrogens (Bocchinfuso et al., 2000; Curtis Hewitt et al., 2000).

Similar to the mouse, the initiating event in human breast development is the formation of a bilateral ridge, or milk line, in the second trimester, which develops into the breast bud that is separated from the surrounding mesenchyme by the basement membrane (Howard and Gusterson, 2000). By 28 weeks, two distinct cell populations can be distinguished in the mammary anlage: a cytokeratin 14 (KRT14)-negative central cell population and a KRT14- and α-smooth muscle actin (SMA, ACTA2)-positive peripheral or basal population (Anbazhagan et al., 1998; Howard and Gusterson, 2000). In the later stages of embryonic development, a rudimentary ductal tree arises from the growing bud as the lumen enlarges and secondary branches emerge, which in turn become canalised, branch and invade the surrounding mesenchyme. In contrast to murine development, human mammary morphogenesis is indistinguishable in females and males in utero (Howard and Gusterson, 2000). A further difference to the mouse is the secretory activity of the human breast and the production of milk in the late-term human foetal and newborn breast in response to
maternal hormones, which is followed by menopause-like involution upon hormone withdrawal (Anbazhagan et al., 1991; Howard and Gusterson, 2000).

1.3.2 Regulators of early mammary gland development and breast cancer

Embryonic development of the mammary gland is largely determined by mesenchymal-epithelial interactions, as mentioned above. Knockout studies in mice have provided valuable insight into the signalling pathways that regulate cell fate decisions in the developing gland (Mikkola and Millar, 2006). Genes that play an important role in the formation of the mammary anlagen and later embryonic development include Wnt10b and its target Lef-1, which are the earliest known markers of placode formation (Howard and Ashworth, 2006). Their expression in turn is dependent on Tbx3, an important regulator of placode induction. Tbx3 was also proposed to function in the specification of mammary cell identity (Davenport et al., 2003; Howard and Ashworth, 2006). The observations made in mice are confirmed by human genetic disorders that are characterised by impaired breast development such as ulnar-mammary syndrome, which is caused by mutations in the TBX3 gene (Bamshad et al., 1997). Further signalling molecules implicated in mammary morphogenesis include the fibroblast growth factor 10 (Fgf10), its receptor Fgfr2 and the epidermal growth factor receptor ligand neuregulin 3 (Nrg3) (Howard and Ashworth, 2006; Howard et al., 2005). Various other signalling factors regulate breast development, which has been reviewed elsewhere (Howard and Gusterson, 2000).

As reassignment of cell identity plays an important role in oncogenic transformation, it is conceivable that parallels exist between the latter and normal mammary development. For instance, variant single nucleotide polymorphisms in the FGFR2 gene are associated with an increased risk of breast cancer, as discussed below. The intricate relationship of mammary development, differentiation of the adult mammary epithelium and tumourigenesis is exemplified by the zinc finger transcription factor GATA-3, deletion of which leads to profound defects in early murine development and perturbation of the maintenance of luminal cell differentiation in the adult gland (Asselin-Labat et al., 2007).
1.3.3 Structure and homeostasis of the adult mammary gland

After birth, development of the human breast remains restricted to isometric growth in both females and males until the onset of puberty when the mammary epithelium and stroma undergo dramatic changes. This is also seen in rodents. In humans, there is an increase in fibrous and adipose tissue in the stroma, which in turn constitutes 80% of the adult nonlactating breast tissue. Proliferation of the connective tissue is followed by expansion of the epithelium in response to estrogen signalling in the female breast at the onset of puberty, while the male breast does not undergo further growth. In the human breast, the primary ducts that reach the nipple elongate and branch dichotomously and sympodially, giving rise to a branched network of increasingly smaller ducts, which in turn end in terminal ducts that give rise to blind-ended ductules called acini. Peripubertal proliferation of the epithelium occurs predominantly at end bud-like structures. At the tip of the extending duct, the terminal ductal lobular unit (TDLU) develops, which consists of a collection of acini embedded in a concentric layer of fibroblasts and which is considered the mature functional unit of the adult breast, while the ducts are surrounded by loose cellular intralobular stroma (Howard and Gusterson, 2000). The TDLUs are also thought to be the predominant site of breast cancer formation (Visvader, 2009). Due to the unavailability of tissue samples, little is known about the morphological changes during pregnancy and lactation that occur in response to systemic hormones in the human breast. Limited data confirm similarities to the events seen in mice including dramatic proliferative activity of the acinar cells, an increased number of lobules and a loss of adipose tissue due to the depletion of fat in adipocytes. At weaning, the mammary epithelium undergoes involution, which is characterised by the removal of secretory epithelial cells by apoptosis and phagocytosis. The epithelial layers of the resting breasts reform in the area of the acini, while the ducts are not affected by expansion and involution for the most part. This is in contrast to menopause-associated involution, when both lobules and ducts are reduced in number and the intralobular fibroblastic stroma is replaced by fat (Howard and Gusterson, 2000). Knockout and transplantation studies have permitted us to gain a detailed understanding of the development of the adult mouse mammary gland. Like in humans, peripubertal development of the murine mammary gland occurs in response to estrogen-, amphiregulin-, epidermal growth factor (Egf)- and insulin-like growth
factor 1 (Igf1)-dependent signalling, resulting in the elongation and bifurcation of the rudimentary ductal tree at its bulbous tip structures termed terminal end buds (TEB) (Sternlicht, 2006). The mitogenic activity of estrogen, which is thought to be the key regulator of development during puberty, is mediated by amphiregulin-dependent induction of the epidermal growth factor receptor (Egfr) in a paracrine manner (Ciarloni et al., 2007). Once the margins of the mammary fat pad are reached in the sexually mature female, ductal branching is refined through tertiary sidebranching in response to cyclic ovarian stimulation through progesterone, which is thought to act through paracrine activation of Wnt4 and Rank ligand (RankL) (Brisken et al., 2000). Finally, during pregnancy, lactogenic differentiation, alveologenesis and further sidebranching occur upon prolactin-dependent stimulation (Brisken and O'Malley, 2010; Sternlicht, 2006). Unlike the human breast, the adult murine ductal tree does not end in TDLUs but the lobules are composed of alveolar buds at the site of regressed TEBs, which form during each oestrous cycle. The alveoli also contain the luminal secretory cells that undergo functional differentiation during pregnancy and lactation. Another important difference to the human breast is that the murine mammary stroma contains significantly less connective and more adipose tissue (Visvader, 2009).

1.3.4 The stem cell hierarchy in the mammary gland

Despite morphological differences between the human and murine mammary gland, the mouse has proved an invaluable model system for normal mammary gland biology as well as breast tumourigenesis. In recent years, there has been particular focus on the delineation of the cellular hierarchy in the mammary gland as the tremendous proliferative capacity and plasticity of the breast inevitably raise the question of the existence of a somatic mammary stem cell.

The normal adult mammary duct has a bi-layered architecture: the lumen of the duct is lined by a single layer of luminal epithelial cells, while myoepithelial cells form a unicellular layer covering the luminal cell layer. The myoepithelial cells are in contact with the basement membrane, a specialised extracellular matrix that separates the epithelial cells from the surrounding stroma that consists predominantly of fibroblasts, adipocytes, endothelial cells and lymphocytes (Fig. 1).
Luminal and myoepithelial cells express a characteristic subset of epithelial cytokeratins and other markers, which are used for diagnostic purposes. Human myoepithelial cells express cytokeratins 5 (K5) and 14 (K14), which are markers for stratified and glandular epithelial cells that are located along the basement membrane (and which are therefore also referred to as basal markers) (Fig. 1). In contrast,
luminal epithelial cells express K18, along with its partnering intermediary filament K8 (Fig. 1). K19 is sometimes used as a luminal marker but its expression is not restricted to the luminal cell fraction and has been proposed to be a “neutral” switch cytokeratin that marks common progenitor cells (Stasiak et al., 1989; Taylor-Papadimitriou et al., 1989; Villadsen et al., 2007). Other markers that have traditionally been used for myoepithelial cells include α smooth muscle actin (SMA, ACTA2) (Fig. 1), neutral endopeptidase CD10, and p63, while human luminal cells are positive for mucin 1 (MUC1) and epithelial cell adhesion molecule (EPCAM) (O'Hare et al., 1991; Stingl et al., 2001).

In analogy to the hierarchical differentiation process that has been established for the haematopoietic system, there is mounting evidence for the existence of a similar differentiation hierarchy in the adult mammary gland. Mammary stem cells (MaSCs) are thought to give rise to all cells of the luminal and myoepithelial lineages via a series of increasingly lineage-restricted intermediate cell types (Fig. 3). It long proved difficult to define, let alone isolate mammary stem cells (MaSC) owing to a lack of adequate assays that would sufficiently confirm the “stemness” of the cell in question. The development of in vivo mammary reconstitution assays in the mouse and the optimisation of tissue dissociation methods over the last 50 years have allowed for the recent prospective isolation of mammary stem cells.

In mice, mammary stem cells are functionally defined as cells that are capable of reconstituting a functional mammary gland in a mammary fat pad where the rudimentary ductal tree has been surgically removed in female pre-pubertal mice, an assay pioneered in 1959 by DeOme and colleagues (DeOme et al., 1959). Removal of the epithelium leaves the “cleared fat pad”, the stromal tissue that consists of adipocytes, fibroblasts, macrophages and endothelial cells (Neville et al., 1998). Classic studies went on to show that mammary epithelial outgrowths could be generated from both small explants and cell suspensions. The site from which explants were taken had no effect on their ability to reconstitute the gland, suggesting that repopulating cells are found throughout the mammary gland. In addition, these cells could be serially transplanted several times, suggesting an ability to self-renew. Furthermore, the capacity of these cells to repopulate the fat pad was shown to be independent of the reproductive history or developmental stage of the donor gland (Daniel et al., 2009; Daniel and Young, 1971; Hoshino and Gardner, 1967; Smith,
1996; Smith and Medina, 1988). Smith and co-workers further demonstrated the existence of different types of progenitor cells that gave rise to ductal or secretory (lobular) outgrowths (Smith, 1996). Using stably MMTV-infected donor cells, they also showed the ability of a single cell to repopulate the entire mammary gland (Kordon and Smith, 1998). They further proposed small light cells (SLC) that exhibited a characteristic lack of cytoplasmic differentiation to be potential mammary stem or progenitor cells (Chepko and Smith, 1997). While these classic studies demonstrated the existence of mammary cells that have the ability to divide asymmetrically (thereby giving rise to the luminal and myoepithelial cell lineages) and self-renew, which are the two characteristics that are thought to define a stem cell, they failed to isolate purified mammary stem cells.

In recent years, improved tissue dissociation techniques in combination with the application of fluorescence-activated cell sorting (FACS) have led to the identification and purification of distinct mammary epithelial cell subpopulations and the prospective isolation of mouse mammary stem cells. Visvader, Eaves, Stingl, Smalley and co-workers first demonstrated the isolation of a discrete mouse mammary stem cell-enriched population based on the expression of a combination of cell surface markers (Shackleton et al., 2006; Sleeman et al., 2006; Stingl et al., 2006). Murine MaSC were found to be highly enriched in the mammary epithelial subpopulation that is characterised by the marker profile CD49f\textsuperscript{high}/CD29\textsuperscript{high}/CD24\textsuperscript{+/med}/Sca1\textsuperscript{−} (Shackleton et al., 2006; Sleeman et al., 2006; Stingl et al., 2006). These cells are able to self-renew and regenerate a functional mammary gland in a cleared fat pad and retain this ability after serial transplantation. Based on this property, they were termed mammary repopulating units (MRUs). They are distinguished from mammary colony-forming cells (Ma-CFCs) (Stingl et al., 2006). Their name is derived from the colony-forming cell (CFC) assay, which was originally developed to test the ability of haematopoietic progenitor cells to proliferate and differentiate into colonies in a semi-solid medium. The Ma-CFC assay is now commonly used to characterise mammary progenitors that give rise to discrete colonies in low-density adherent culture (Bachelard-Cascales et al., 2010; Stingl et al., 2001; Stingl et al., 2006). Limiting-dilution analysis showed that a single cell (or MRU) had the capability to give rise to the entire ductal tree, impressively demonstrating the multipotency of the proposed stem cell (Shackleton et al., 2006; Stingl et al., 2006). It has to be noted however that these stem cells represent less than
10% of the CD49$^{\text{high}}$/CD29$^{\text{high}}$/CD24$^{+}/\text{med}/\text{Sca1}^-$ subset, which also contains mature myoepithelial cells and probably other basal cell intermediates such as myoepithelial progenitor cells, which remain to be characterised. The similar FACS profiles of MaSC and other myoepithelial/basal cells may reflect their common basal location in the duct. FACS has become the method of choice for the purification and analysis of mammary epithelial subpopulation but when comparing studies from different laboratories, differences in the experimental approach and the antibodies used for cell sorting need to be considered. The CD49$^{\text{high}}$/CD24$^{\text{med}}$ population described by Stingl and co-workers (Stingl et al., 2006) is thus thought to be identical to the CD29$^{\text{high}}$/CD24$^+$ population identified by Shackleton and colleagues (Shackleton et al., 2006) and the CD49$^{\text{high}}$/CD24$^{+}$/CD61$^+$ population described by Sleeman et al. (Sleeman et al., 2006). Furthermore, CD24$^+$ cells express both CD49f (α6 integrin) and CD29 (β1 integrin) at high levels, suggesting a role for these two integrins in the interaction with the extracellular matrix (ECM) in the basal, stem cell-enriched compartment. In keeping with this, deletion of β1 integrin from the basal compartment results in a reduced repopulation capacity (Taddei et al., 2008).

Analogous to the MaSC-enriched basal subpopulation, luminal progenitors and their mature progeny have been described based on their FACS profile. Luminal progenitors were described as CD24$^{+}$/CD29$^{\text{low}}$/CD61$^+$, while mature luminal cells were shown to be CD24$^{+}$/CD29$^{\text{low}}$/CD61$^-$ in the original publications (Shackleton et al., 2006; Sleeman et al., 2006; Stingl et al., 2006). Smalley and colleagues further showed that prominin staining distinguishes two different subpopulations within the luminal group, where prominin $1^+$ cells represent a hormone-sensing luminal subcompartment, while prominin $1^-$ cells are involved in milk production, as they express hormone receptor and milk proteins, respectively. Sca1 staining essentially gave the same results (Sleeman et al., 2007). Recently, Visvader and colleagues identified the tyrosine receptor kinase Kit as an additional marker for luminal progenitor cells (Lim et al., 2009).
Establishing similar stem cell assays for human mammary stem cells has posed a particular challenge due to a lack of equivalent model systems. Similar to the mouse mammary gland, *in vitro* assays in combination with FACS analysis of human mammary epithelial cells have provided evidence for a hierarchical structure of the human breast and identified markers for the different human mammary cell lineages. 2D and 3D *in vitro* culture of dissociated cells or organoids derived from reduction mammoplasties has been useful to determine the capability of individual epithelial subpopulations to give rise to differentiated progeny and thus to define mammary cell lineages (Bachelard-Cascales et al., 2010; Duss et al., 2007; Stingl et al., 1998; Stingl et al., 2001; Visvader, 2009). Magnetic beads coupled to antibodies directed against cell surface markers can be used instead of FACS to sort cells (Proia et al., 2011).
However some argue that, albeit less technically challenging and presumably gentler on the cells, the immunobead method results in a less exact separation of the different epithelial subpopulations. As described above for murine cells, the mammary colony-forming cell assay is used to investigate the ability of distinct human progenitor cells to form colonies in collagen gel-based medium on a layer of irradiated NIH 3T3 mouse fibroblasts (Bachelard-Cascales et al., 2010; Eirew et al., 2008) (Fig. 2). Choosing an alternative approach, Dontu and colleagues established a protocol for the non-adherent culture of human primary mammary cells. The suspension culture of dissociated cells enriches for mammary stem and progenitor cells, which are thought to survive and give rise to spherical structures termed mammospheres. In contrast, the non-adherent culture conditions negatively select for more differentiated cell types that are thought to die from anoikis. Dontu and co-workers demonstrated that single cells derived from dissociated mammospheres were able to differentiate along the luminal and myoepithelial lineages, which was confirmed by the presence of differentiation markers and the expression of basal and luminal keratins, respectively (Dontu et al., 2003). The mammosphere assay has since been employed by several groups to select for and maintain normal mammary stem and progenitor cells as well as tumour-initiating cells (Duss et al., 2007; Harrison et al., 2010). However, more recent findings call the validity of mammosphere culture as a true stem cell assay into question as the self-renewal potential of normal mammary epithelial cells in suspension is exhausted within five passages (Dey et al., 2009 and my own unpublished observation). This may suggest that mammosphere culture allows for the short-term maintenance of common progenitors, which are able give rise to luminal and myoepithelial cells under differentiating conditions but are not able to self-renew. Alternatively, the cell culture conditions may be a limiting factor, allowing true mammary stem cells to give rise to more differentiated progeny but not to self-renew. Indeed, there is mounting evidence that in vitro cell culture conditions have a profound impact on the phenotype of the cultured cells (Ince et al., 2007; Stingl et al., 2001 and my own unpublished observations). This is addressed in more detail below.
Figure 3. Epithelial cell hierarchy in the mammary gland.

Proposed model of the differentiation hierarchy within the mammary epithelium. Cell surface markers commonly used for the isolation of distinct subpopulations of mouse (blue) and human (green) mammary epithelial cells are shown. The cell surface markers are described in Table 1.

In a pioneering study using an orthotopic xenograft model, Kuperwasser and co-workers have shown that human primary mammary cells, like their murine counterparts, have the ability to reconstitute human mammary tissue when injected into the cleared fat pad that was “humanised” by pre-injection of human mammary fibroblasts. As in the mouse, this in vivo assay confirmed the existence of multipotent cells in the human mammary gland (Kuperwasser et al., 2004; Proia and Kuperwasser, 2006). Visvader and co-workers further refined the assay using dissociated single cells instead of organoids (Lim et al., 2009). Dontu and colleagues have used this xenograft assay to show that human mammary stem and progenitor cells in both the normal breast and breast tumours express high levels of aldehyde
Introduction
dehydrogenase 1 (ALDH1) (Ginestier et al., 2007). An alternative non-orthotopic in vivo assay was developed by Eaves and colleagues. They showed that dissociated human epithelial cells give rise to bilayered ductal structures when implanted under the kidney capsule of hormone-treated immunodeficient mice (Eirew et al., 2010; Eirew et al., 2008) (Fig. 2). Bachelard-Cascales and colleagues have further shown, using a 3D cell-based assay, that human MaSC or early progenitors are able to give rise to ductal outgrowths when injected into matrigel (Bachelard-Cascales et al., 2010). This may be a useful alternative to in vivo reconstitution assays as the cell culture-based assay is less labour intensive. Based on the observations from in vitro clonogenic and in vivo reconstitution assays, human mammary stem cells are currently thought to be $CD49f^{\text{high}}/\text{EpCAM}^{+}/\text{CD24}^{+}$ (Eirew et al., 2008; Lim et al., 2009; Stingl et al., 1998; Stingl et al., 2001) as well as ALDH1$^+$ (Ginestier et al., 2007). Human $CD49f^{\text{high}}/\text{EpCAM}^{+}/\text{CD24}^{+}$ cells are able to repopulate the cleared fat pad and have, albeit limited, self-renewal capacity (Eirew et al., 2008; Lim et al., 2009). ALDH1$^+$ cells can also reconstitute the mammary gland but their self-renewal properties have not been defined (Ginestier et al., 2007). Bipotent common progenitors exhibit high levels of CD10 in addition to high levels of CD49f and low levels of EpCAM (Bachelard-Cascales et al., 2010; Stingl et al., 1998; Stingl et al., 2001). CD10$^+$ cells form mammospheres and give rise to luminal and myoepithelial lineages in vitro but as their in vivo repopulating capacity has not been determined, it is unclear whether they are true stem cells or common progenitors (Bachelard-Cascales et al., 2010). Luminal-restricted progenitors are $CD49f^{\text{high}}/\text{EpCAM}^{+}/\text{CD133}^{+}/\text{CD24}^{+}$ (Lim et al., 2009; Stingl et al., 1998; Stingl et al., 2001) and also express high levels of KIT, as recently shown by the Visvader group (Lim et al., 2009). In contrast, mature luminal cells have the phenotype CD49f$^{-}$/EpCAM$^{-}$ and express the classic luminal marker MUC1 but are CD10$^-$ (Stingl et al., 1998; Stingl et al., 2001). No specific myoepithelial progenitor marker has emerged thus far, although this compartment can be distinguished from the luminal lineage by the absence of MUC1 and EpCAM expression and the presence of CD10 expression (Stingl et al., 1998; Stingl et al., 2001). As seen in the mouse, human mature myoepithelial cells share the expression profile of the stem cell compartment (CD49f$^{\text{high}}$/EpCAM$^{-}$). As mentioned, this may reflect their shared basal position in the duct but underlines the requirement for reliable functional stem cell assays to distinguish between phenotypically similar cells. The expression of cell surface
markers suggests that the epithelial hierarchies in the human and murine mammary gland are generally conserved, as reflected for instance by the high CD49f expression that is common to both human and murine mammary stem and progenitor cells. However, not all cell surface markers are conserved in their expression pattern. While the expression of CD24 is restricted to luminal cells in the human breast, it appears to be more widely expressed in the mouse mammary gland, marking all epithelial cells to varying degrees (Lim et al., 2009; Raouf et al., 2008; Shackleton et al., 2006; Sleeman et al., 2006; Stingl et al., 2006).

In addition to the cell types discussed, there is evidence for the existence of an alveolar progenitor cell as well as pregnancy-associated cell types, which are addressed in more extensive reviews (Stingl, 2009; Visvader, 2009).

The cell surface markers commonly used for the purification of mouse and human mammary epithelial cells are detailed in Table 1. For a summary of the current understanding of the epithelial cell hierarchy in the mouse and human mammary glands, see Figure 3. Although the cornerstones have been confirmed, the model remains somewhat speculative at this point, as it is unclear how many different intermediates exist between the stem cell, bipotent progenitor (or transit-amplifying) cell and their most differentiated progeny. Similarly, it is not known at what point exactly the ductal lineage divides into the different specialised ductal cell types. An alternative model to the one depicted may involve individual ductal and alveolar lineages (Smith, 1996).
### Table 1. Cell surface markers used for mammary epithelial cell sorting

<table>
<thead>
<tr>
<th>Marker</th>
<th>Official gene symbol</th>
<th>Name</th>
<th>Biological Function (Reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mouse</strong></td>
<td></td>
<td></td>
<td>Mammary epithelial subpopulation (Reference)</td>
</tr>
<tr>
<td>CD24</td>
<td>CD24</td>
<td>CD24 molecule, heat stable antigen</td>
<td>Glycosyl phosphatidylinositol-anchored glycoprotein, expressed on a variety of cell types including granulocytes and B cells (Fang et al., 2010). CD24&lt;sup&gt;high&lt;/sup&gt;: luminal, MaSC-enriched, myoepithelial progenitor CD24&lt;sup&gt;low&lt;/sup&gt;: myoepithelial (Shackleton et al., 2006; Sleeman et al., 2006).</td>
</tr>
<tr>
<td>CD29</td>
<td>Itgb1</td>
<td>Integrin β1</td>
<td>Adhesion receptor, binds to various ECM components (collagen, fibronectin, laminin) upon heterodimerisation with different α integrins. Expressed on leukocytes, endothelial cells and epithelial cells (Hynes, 2002; Plow et al., 2000; Visvader, 2009). CD29&lt;sup&gt;high&lt;/sup&gt;: MaSC CD29&lt;sup&gt;low&lt;/sup&gt;: luminal progenitor, mature luminal (Shackleton et al., 2006).</td>
</tr>
<tr>
<td>CD49f</td>
<td>Itga6</td>
<td>Integrin α6</td>
<td>Adhesion receptor, heterodimers bind laminin expressed ubiquitously. Expressed on T cells, monocytes, endothelial cells and epithelial cells (Hynes, 2002; Plow et al., 2000). CD49f&lt;sup&gt;high&lt;/sup&gt;: MaSC, myoepithelial CD49f&lt;sup&gt;low&lt;/sup&gt;: luminal (Sleeman et al., 2006; Stingl et al., 2006).</td>
</tr>
<tr>
<td>CD61</td>
<td>Itgb3</td>
<td>Integrin β3</td>
<td>Adhesion receptor, heterodimers bind RGD-containing proteins, expressed on endothelial cells and angiogenic vascular tissue (Hynes, 2002; Plow et al., 2000). CD61&lt;sup&gt;+&lt;/sup&gt;: luminal progenitor, cancer initiating cell CD61&lt;sup&gt;+&lt;/sup&gt;: mature luminal (Shackleton et al., 2006; Vaillant et al., 2008).</td>
</tr>
<tr>
<td>CD133</td>
<td>PROM1</td>
<td>Prominin 1</td>
<td>Pentaspan transmembrane glycoprotein, expressed on adult stem cells, thought to contribute to stem cell maintenance by inhibiting differentiation (Meregalli et al., 2010). Prominin 1&lt;sup&gt;+&lt;/sup&gt;: mature ER&lt;sup&gt;+&lt;/sup&gt; luminal Prominin 1&lt;sup&gt;+&lt;/sup&gt;: luminal progenitor, mature ER&lt;sup&gt;+&lt;/sup&gt; luminal, alveolar (Sleeman et al., 2007).</td>
</tr>
<tr>
<td>Sca1</td>
<td>Ly6a</td>
<td>Lymphocyte antigen 6 complex, stem cell antigen 1</td>
<td>Phosphatidylinositol-anchored protein, expressed on subpopulations of bone marrow, T lymphocytes, B lymphocytes, early thymic cells and non-haematopoietic tissues (Visvader, 2009). Sca1&lt;sup&gt;+&lt;/sup&gt;: ER&lt;sup&gt;+&lt;/sup&gt; luminal (Shackleton et al., 2006; Sleeman et al., 2007); progenitor? (Welm et al., 2002) Sca1&lt;sup&gt;+&lt;/sup&gt;: MaSC (Shackleton et al., 2006).</td>
</tr>
<tr>
<td>CD31</td>
<td>Pecam 1</td>
<td>Platelet/endothelial cell adhesion molecule 1</td>
<td>Endothelial marker CD31 and haematopoietic markers CD45 and Ter119 taken together are termed lineage (Lin); used to pre-sort mammary cells in order to deplete the heterogeneous mammary cell mixture from these cells, leaving the epithelial compartment (Lin&lt;sup&gt;-&lt;/sup&gt;) (Shackleton et al., 2006).</td>
</tr>
<tr>
<td>CD45</td>
<td>Ptprc</td>
<td>Protein tyrosine phosphatase, receptor type, C</td>
<td></td>
</tr>
<tr>
<td>TER119</td>
<td>Ly76</td>
<td>Lymphocyte antigen 76</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. continued

<table>
<thead>
<tr>
<th>Marker</th>
<th>Official gene symbol</th>
<th>Name</th>
<th>Biological Function (Reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mammary epithelial subpopulation (Reference)</td>
</tr>
</tbody>
</table>
| Human  |                     |                               | Cytosolic member of the aldehyde dehydrogenase family, functions as detoxifying enzyme responsible for the oxidation of intracellular aldehydes. Role in haematopoietic stem cell differentiation by oxidizing retinol to retinoic acid (Chute et al., 2006).
| ALDH1  | ALDH1A1             | Aldehyde dehydrogenase 1 family, member A1 | ALDH1^{+}: MaSC, cancer stem cells (Ginestier et al., 2007) |
|        |                     | Membrane metallo-endopeptidase | CD10^{+}: stem cells and early progenitors (Bachelard-Cascales et al.); myoepithelial cells (O’Hare et al., 1991; Shipitsin et al., 2007; Stingl et al., 1998); myoepithelial progenitor (Stingl et al., 2001) |
| CD10 CALLA | MME                     | Membrane metallo-endopeptidase | CD10^{+}: luminal progenitor (Stingl et al., 2001) |
|        |                     |                               | CD10 variable: common progenitor (Stingl et al., 2001) |
| CD24   | CD24               | CD24 molecule                  | CD24^{+}: luminal (Shipitsin et al., 2007), luminal progenitors (Lim et al., 2009) |
|        |                     |                               | CD24^{low}: tumour-initiating cells (Al-Hajj et al., 2003; Ponti et al., 2005), MaSC (Lim et al., 2009) |
| CD29   | ITGB1              | Integrin β1                   | CD29^{high}: MaSC and myoepithelial cell (Bachelard-Cascales et al., 2010) |
| CD44   | CD44               | CD44 molecule                  | CD44^{+}: E- and L-selectin ligand; heparan sulfate proteoglycan |
|        |                     |                               | CD44^{−}: tumour-initiating cells (Al-Hajj et al., 2003; Ponti et al., 2005); progenitors (Shipitsin et al., 2007) |
| CD49d  | ITGA4              | Integrin α4                   | CD49d^{high}: myoepithelial (E. Bachelard-Cascales, personal communication) |
|        |                     |                               | Function see above. |
| CD49f  | ITGA6              | Integrin α6                   | CD49f^{high}: common progenitors, MaSC, myoepithelial |
|        |                     |                               | CD49f^{low}: mature luminal (Bachelard-Cascales et al., 2010; Eirew et al., 2008; Lim et al., 2009; Stingl et al., 2001) |
| CD90   | THY1               | Thy-1 cell surface antigen    | Heavily glycosylated glycoprophatidylinositol-anchored member of the immunoglobin superfamily. Expressed on haematopoietic stem cells, thymocytes and neurons. Role in adhesion and signal transduction in T cells (Visvader, 2009). |
|        |                     |                               | CD90^{+}: Bipotent progenitor, myoepithelial differentiated (Raouf et al., 2008) |
| CD133  | PROM1              | Prominin 1                    | Function see above. |
| EpCAM ESA | EPCAM              | Epithelial cell adhesion molecule | EpCAM^{+}: luminal progenitor, mature luminal |
|        |                     |                               | EpCAM: myoepithelial progenitor, mature myoepithelial (Stingl et al., 2001; Stingl et al., 2006) |
Like most cell types, mammary stem cell function is dependent on the interactions between the stem cell and stromal components constituting its environment, which is referred to as the stem cell niche. The local environment surrounding the stem cell is in turn influenced by systemic hormonal cues as the breast undergoes developmental and cyclic changes (reviewed in Brisken and Duss, 2007). The requirement of interactions between the mammary stem cell and its stromal environment itself may pose a problem for the isolation and subsequent characterisation of the mammary stem cell as the available purification techniques may compromise the viability and functionality of the stem cell merely by removing the environment that allows for maintenance of “stemness”. Mammary gland reconstitution from a single cell has shown convincingly that interactions between mammary stem cells and neighbouring epithelial cells are not required for repopulation activity of the stem cell (Shackleton et al., 2006), suggesting a role for the stromal environment in the maintenance of stem cell function. However, Smith and colleagues recently demonstrated that normal mammary epithelial cells are able to redirect malignant MMTV-neu tumour cells to form normal mammary structures (Booth et al., 2011), refuting the notion that neighbouring epithelial cells do not contribute to the stem cell niche. It has long been known that stromal cues are important for the early development of the mammary gland (Howard and Gusterson, 2000). Studies by Smith and co-workers have shown that in the mouse, the mammary stroma is able to reprogram stem cells from other organs, such as neural stem cells and testicular cells, to adopt a mammary epithelial cell fate during gland regeneration (Booth et al., 2008; Boulanger et al., 2007). Notably, the murine mammary microenvironment is also capable of redirecting...
human embryonal carcinoma cells to produce differentiated mammary epithelial cell progeny in xenografts (Bussard et al., 2010). This is consistent with a study by Bissel and colleagues showing instructive effects of the human extracellular matrix on cell fate determination (LaBarge et al., 2009). The capacity of murine cells to reprogramme human epithelial cell fate calls into question the need for a “humanisation” of the murine mammary gland prior to xenografting human cells (Kuperwasser et al., 2004). The stromal components that constitute the mammary stem cell niche are yet to be identified. Macrophages may play a role because the regenerative capacity of MaSC is reduced in the mouse fat pad that has been depleted of macrophages but the signalling pathways involved remain to be identified (Gyorki et al., 2009). Further evidence that the stromal microenvironment is essential to the adequate functioning of MaSC or early progenitors comes from in vitro assays. Human progenitors require the presence of feeder cells in order to produce differentiated progeny, and in the 3D in vitro TDLU assay, single FACS-sorted cells are able to give rise to ductal outgrowths in matrigel only when supplemented with fibroblasts (Bachelard-Cascales et al., 2010; Eirew et al., 2008). In the mouse, MaSC activity is predominantly found in the region of the peripheral cap cells of terminal end buds (TEBs). As mentioned above, in the human breast, TEBs are not a prominent structure. Human MaSC are thought to be located in the ducts, rather than in the terminal ductal lobular units (TDLUs) (LaBarge et al., 2007; Villadsen et al., 2007). As the developing gland expands and the location of the TLDUs changes, it appears likely that MaSC present in mature ducts are silenced by stromal signals. Similarly, during periods where MaSC activity is not required, for instance at involution, MaSC activity may be inhibited. One such candidate negative regulator of MaSC proliferation is TGFβ1, which appears to act via stromal intermediates (Boulanger et al., 2005; Pierce et al., 1993; Silberstein et al., 1992). The existing data suggest an important function for the stromal microenvironment in MaSC homeostasis. However, hormonal control of MaSC proliferation appears to be predominantly mediated by the mammary epithelium, as discussed below.

Currently available methods have provided great insight into the hierarchical structure of the mammary gland but the isolation, dissociation and culture of primary epithelial cells remain a challenge as they may significantly alter the phenotype of the cells of interest. The truly physiological assay to delineate the hierarchy and location of
mammary epithelial subtypes in the normal breast and to identify the cell of origin of the different tumour subtypes will be lineage-tracing of individual mammary cells, which was elegantly shown for stem cells in the colon by Clevers and co-workers using the stem cell marker Lgr5 (Barker et al., 2007).

1.3.5 Regulation of mammary gland development and homeostasis by steroid hormone receptors

The importance of hormonal signalling for the development of the mammary gland is well established. While most of the early development is dependent on epithelial-mesenchymal interactions, hormonal control of mammary gland development begins with the onset of puberty. In both humans and mice, pubertal development is controlled by hormones that are released by the hypothalamus, ovaries and pituitary. In the human, puberty begins with the first release of gonadotropin-releasing hormone (GnRH) from the hypothalamus. GnRH stimulates the release of leutinising hormone (LH) and follicle stimulation hormone (FSH) from the pituitary, which in turn induce the maturation of the ovarian follicles and the subsequent release of estrogen and progesterone from the corpus luteum (Jones and Lopez, 2006). In the mouse, ovarian estrogen and growth hormone (GH) stimulation leads to the elongation and bifurcation of the growing ductal tree in the first weeks of puberty by inducing a rapid increase in proliferation at the terminal end buds, followed by secondary and tertiary side branching in sexually mature females, which require ovarian progesterone signalling (Brisken and O'Malley, 2010). Ductal morphogenesis is thus predominantly controlled by estrogen and progesterone signalling, whereas alveolar morphogenesis during pregnancy is regulated by prolactin and progesterone (Oakes et al., 2008b). 17-β-estradiol (E2) acts directly on the normal mammary epithelium to induce proliferation (Daniel et al., 1987; Silberstein et al., 1994). In addition, it activates downstream mammogenic hormone signalling, for instance by inducing progesterone receptor expression (Shyamala et al., 1990). Conversely, ERα knockout mice fail to undergo normal breast development and possess only a rudimental ductal tree (Bocchinfuso and Korach, 1997; Korach et al., 1996; Mallepell et al., 2006). ERα has thus emerged as a master regulator of mammary gland development but there are conflicting views as to the identity of the ERα-expressing epithelial cell in the normal
breast, as discussed below. Like in normal development, 17-β-estradiol has a mitogenic effect in hormone-sensitive breast cancer cells. As mentioned above, about 70% of all breast cancers are defined by the expression of ERα, and there is evidence that cumulative exposure to high estrogen levels increases breast cancer risk (Andersen and Poulsen, 1989; Anderson et al., 1998; Yager and Davidson, 2006). At the end of the 1980s it became clear that ERα activates the transcription of the MYC and cyclin D1 (CCND1) genes in response to 17-β-estradiol, thereby promoting entry into S phase of the cell cycle and inducing proliferation in cell cycle-arrested breast cancer cells (Altucci et al., 1996; Dubik et al., 1987; Dubik and Shiu, 1988; Prall et al., 1998). Accordingly, successful anti-estrogen therapy in patients with ERα-positive breast cancer demonstrates that estrogen signalling is required for the proliferation of ERα-positive (ER+) tumour cells. Paradoxically, there appears to exist a crucial difference between ERα-dependent signalling in the normal breast and in breast cancer. ER+ breast tumour cells themselves proliferate. This is in contrast to the situation in the normal breast. It has long been known that only a subset of cells in the mammary epithelium expresses ERα but their identity has been subject to controversy. Even within the luminal population in the mouse, only about 5% of CD61+ luminal progenitor cells and about 30-40% of the mature luminal cells express the estrogen receptor. Several groups proposed that putative mammary stem cells express ERα, which would correspond to the autocrine mitogenic function of estrogen in breast cancer (Booth and Smith, 2006; Cheng et al., 2004; Clarke et al., 2005). However, more recent data demonstrate that MaSC themselves do not express ERα but are stimulated by neighbouring ERα+ cells in a paracrine manner (Asselin-Labat et al., 2006; Lim et al., 2009; Mallepell et al., 2006). To address this question, Smalley and co-workers prospectively isolated ER+ and ER− cells from the mouse using the surface markers CD24, Prominin-1 and Sca-1. They showed that cells that are CD24high/Sca-1+/Prominin-1+ are also ER+, express the progesterone and prolactin receptors and show higher expression of luminal keratins than ER− luminal cells, which have the phenotype CD24high/Sca-1−/Prominin-1−. ER− luminal cells showed higher expression of the milk proteins lactotransferrin and β-casein, indicating alveolar differentiation, and were enriched for mammary colony-forming cells. Importantly, neither ER−, nor ER− luminal subpopulations produced outgrowths when transplanted in the cleared fat pad, while only the CD24low/Sca-1−/Prominin-1−
myoepithelial subpopulation was enriched for mammary repopulating cells (Sleeman et al., 2007). This study provided direct proof that MaSC in the mouse mammary gland are not ER$^+$ but rather that the ER$^+/PR^+/PrlR^+$ cells are part of a hormone-sensing epithelial subcompartment that relays a mitogenic signal to neighbouring MaSC and progenitors in a paracrine manner. The mechanism of paracrine stimulation of MaSC remains poorly understood at this point. One candidate for a mediator of estrogen signalling in the normal breast is the EGFR ligand and ER$\alpha$ transcriptional target amphiregulin (Ciarloni et al., 2007). A recent study proposes FGF, its receptor FGFR and the transcription factor TBX3 as mediators of estrogen signalling-dependent expansion of cancer-initiating cells in a paracrine fashion (Fillmore et al., 2010).

It appears that estrogen is not the only hormone that exerts its proliferative effect in an indirect manner. Several recent studies have shown that progesterone signalling induces proliferation of MaSC as well as tumour cells in a paracrine fashion that is mediated by the osteoclast differentiation factor RANKL (Asselin-Labat et al., 2010; Beleut et al., 2010; Fernandez-Gonzalez et al., 2010; Joshi et al., 2010; Schramek et al., 2010).

Although the underlying mechanisms are still not well understood, it appears clear at this point that estrogen signalling has opposite effects in the normal mammary gland and during tumourigenesis. In the normal gland, ER$^+$ cells do not proliferate themselves but merely act as sensors that relay the mitogenic signal to MaSC and, presumably, progenitors or transit-amplifying cells. Moreover, forced expression of ER$\alpha$ in normal human mammary epithelial cells rapidly leads to their growth arrest and differentiation (Duss et al., 2007 and my own unpublished observation). This is in stark contrast to the situation seen in primary luminal breast tumours and breast cancer cell lines, where ER$^+$ expression and estrogen signalling are directly correlated with the proliferative state of the cancer cell. This suggests that during the development of luminal breast cancer, ER$^+$ cells gain the ability to proliferate by inhibiting the differentiating and anti-proliferative effects of estrogen signalling, presumably through the inappropriate expression of a regulator of ER-dependent signalling. It is therefore crucial to identify genes that co-regulate or modulate ER-dependent transcriptional programmes, as resolving this paradox is central to the understanding of the origin of luminal breast cancer.
1.3.6 Regulators of stem cell activity, lineage commitment and differentiation

Much effort has been put into the identification and isolation of prospective mammary stem cells, yet little is known about the factors that regulate cell fate decisions in the mammary gland. Perturbation of the signalling pathways that determine self-renewal of stem cells and differentiation of progenitor cells along a particular lineage likely results in oncogenic transformation. Insight into the regulation of mammary lineage specification comes from expression profiling of the individual epithelial subpopulations in the normal mammary gland as well as of the different breast cancer subtypes (Kendrick et al., 2008; Kouros-Mehr and Werb, 2006; Lim et al., 2010; Raouf et al., 2008). The information obtained from these studies is vast but knowledge of transcriptional co-regulation of genes in a certain subpopulation or tumour subgroup alone does not necessarily help to discern the genes that drive lineage commitment from the genes that are merely co-expressed and play a minor role in the respective cell type. A valuable method to validate candidate regulators of stem cell activity, lineage commitment and differentiation is viral-based stable expression of the gene of interest in normal mammary epithelial cells, followed by characterisation of the resulting phenotype of the transduced cells in transplants or in culture. Amongst the genes identified by expression microarrays are a myriad of transcription factors. For instance, Sox6 was identified exclusively in murine ER-luminal cells, and ectopic expression of Sox6 was shown to promote luminal differentiation (Kendrick et al., 2008). Similarly, knockdown of the Notch target Cbf-1 results in increased MaSC proliferation, while expression of Notch commits MaSC to differentiate exclusively along the luminal lineage (Bouras et al., 2008). Other transcription factors that are involved in the regulation of stem cell renewal, lineage commitment and terminally differentiated cell function in the adult mammary gland include GATA3, FOXA1, ELF5, WNT and TBX3 (Asselin-Labat et al., 2007; Brisken et al., 2000; Fillmore et al., 2010; Kouros-Mehr et al., 2006; Lupien et al., 2008; Oakes et al., 2008a). Considering their role in the regulation of cell fate in the adult mammary gland, it is not surprising that many of these transcription factors also have a function in embryonic development. Conversely, aberrant expression of these transcriptional regulators leads to oncogenic transformation. This general pattern is exemplified by the transcription factor GATA3, which is required for embryonic
development of the mammary gland and which acts as a luminal differentiation factor in the adult gland (Asselin-Labat et al., 2007; Kouros-Mehr et al., 2008). Conversely, GATA3 deficiency marks a loss of luminal differentiation and is linked to tumour progression (Kouros-Mehr et al., 2008). Similarly, the Forkhead transcription factor FOXA1 is a marker of differentiated luminal cells and has been implicated in the control of mammary epithelial cell fate. FOXA1 is thought to act as a pioneer factor in the recruitment of ERα to promoter elements and subsequent transcriptional induction. Using chromosome-wide binding site mapping, Brown and colleagues have demonstrated that FOXA1 regulates ERα-dependent transcription in a lineage-specific manner in estrogen-dependent MCF-7 breast cancer cells versus androgen-dependent LNCaP prostate cancer cells (Carroll et al., 2005; Lupien et al., 2008). Like the differentiation of mammary epithelial cells, stem cell self-renewal and proliferation are controlled by transcriptional regulators, and factors promoting or repressing stem cell properties are attractive candidates for oncogenes and tumour suppressor genes, respectively. Notably, a number of transcription factor genes, including the FOXA1-related Forkhead protein FOXC1, appear to be differentially methylated depending on mammary epithelial cell type and state of differentiation (Bloushtain-Qimron et al., 2008). The polycomb protein BMI1 has emerged as an important regulator of both MaSC activity and luminal differentiation (Pietersen et al., 2008). BMI1 represses p53- and retinoblastoma protein (RB)-induced senescence by silencing p14ARF and p16INK4A expression and at the same time inhibiting MYC-dependent apoptosis (Jacobs et al., 1999a; Jacobs et al., 1999b). Wicha, Dontu and colleagues demonstrated that BMI1 is a downstream target of the hedgehog, Notch and Wnt signalling pathways. Its stem cell self-renewal promoting properties were measured as an increase in mammosphere-initiating cells (Liu et al., 2006). Like other transcriptional regulators that are involved in the control of lineage commitment in the mammary gland, members of the polycomb family including BMI1 are overexpressed in breast cancer, and BMI1 in particular is overexpressed in ER+ breast tumours (Duss et al., 2007; Kim et al., 2004; Sparmann and van Lohuizen, 2006). The study of transcriptional regulators in the mammary gland thus provides valuable insight into the pathways implicated in breast development and differentiation as well as oncogenesis.
1.4 Breast cancer subtypes and the cell of origin

As discussed above, human breast cancers are heterogeneous, both in their molecular profiles and pathology. One interpretation of the remarkable differences found between the individual breast cancer subtypes is that tumours can originate in different mammary epithelial cell types, either in mammary stem cells or in their progeny, at any level of differentiation. Alternatively, cancer may arise in a restricted set of more primitive cells that have a differing potential to differentiate. The delineation of the epithelial cell hierarchy in the normal mammary gland is thus intimately linked to the identification of the cell of breast cancer origin. In addition to this inter-tumoural heterogeneity, differences are found within a patient’s tumour. This intra-tumoural heterogeneity is exemplified by variable ER expression levels within a single tumour (Allred et al., 2004).

Two models of cancer progression, which both account for tumour heterogeneity, have been proposed. For many years, the clonal evolution model of breast tumourigenesis predominated, according to which a normal epithelial cell over time acquires a number of driver mutations, which promote oncogenic transformation of that cell. In this model, all cells in a tumour originally have a similar capacity for tumour propagation and dissemination, and intra-tumoural heterogeneity results from the evolution of distinct clones following a series of mutations.

In contrast, the cancer stem cell hypothesis proposes a hierarchical model, whereby normal mammary stem cells or early progenitor cells are the target of transforming events, which convert them into the cell of cancer origin. In analogy to the model of hierarchical differentiation in the normal mammary gland, these cancer stem cells (CSC) or tumour-initiating cells give rise to tumour cell progeny, thereby promoting tumour formation. The cancer stem cell hypothesis is based on the observation that a subset of cells in primary breast tumours exhibit similarities to stem cells in adult tissues, including the abilities to self-renew and to give rise to heterogeneous and differentiated progeny. A combination of the two models is likely to reflect the actual order of events that lead to tumour formation. It is conceivable that tumour-initiating cells arise from differentiated epithelial cells that acquire one or more mutations causing them to revert back to a less differentiated state and to acquire stem cell properties. Alternatively, adult MaSCs themselves may be the target of oncogenic
transformation, thereby gaining independence from external regulatory cues while at
the same time retaining their original ability to self-renew and to give rise to quickly
dividing progeny. It also appears possible that early progenitor cells are subject to a
block to differentiation at a certain developmental stage or that they undergo aberrant
differentiation. Evidence for the cancer stem cell model comes from studies
demonstrating that only a subset of primary breast tumour cells exhibits tumour-
initiating properties in vivo and in vitro. These stem-like cancer cells were isolated
using markers that had previously been used to purify putative adult stem cells from
various organs (Table 1). Putative human breast tumour-initiating cells with the
phenotype CD44+/CD24−low and ALDH1+ give rise to mammospheres under non-
adherent culture conditions and are able to form tumours when injected into the
mammary fat pad of immunodeficient mice (Al-Hajj et al., 2003; Ginestier et al.,
2007; Grimshaw et al., 2008; Ponti et al., 2005). The latter in vivo approach is
arguably the most meaningful CSC assay. Using expression and DNA methylation
profiling, Polyak and colleagues further demonstrated the expression of known stem
cell markers in the CD44+ subset, which was correlated with decreased patient
survival (Bloushtain-Qimron et al., 2008; Shipitsin et al., 2007). This is consistent
with a study by Pece and colleagues, who demonstrated by labelling non-dividing
putative CSC with the lipophilic dye PKH26 that poorly differentiated, higher grade
tumours are enriched in mammosphere-forming and tumour-initiating cells (Pece et
al., 2010). Putative mammary cancer stem cells have also been identified in mouse
models of cancer using various stem cell surface markers. Tumour-forming CSC were
thus described in MMTV-wnt-1 mice as Thy1+/CD24+ as well as CD61+ (Cho et al.,
2008; Vaillant et al., 2008). The CD44+/CD24−low phenotype has further been
employed to isolate putative tumour-initiating cells in established breast cancer lines
such as MCF-7 and MDA-MB-231. A subset of these cells indeed express the stem
cell marker combination and were shown to generate mammospheres at a higher
frequency than their CD44− and CD24high counterparts but their tumourigenicity in vivo has not been examined (Harrison et al., 2010; Ponti et al., 2005). It appears
somewhat surprising that, after long-term in vitro culture of established carcinoma
cell lines, tumour-initiating cells with stem cell properties are contained in what was
traditionally presumed to be a rather homogenous population of more or less
differentiated tumour cells. On the other hand, if what distinguishes cancer-initiating
cells from their non-tumourigenic counterparts in heterogeneous tumour tissue is their
“stemness”, it is plausible that established breast cancer cell lines are constantly replenished from a small pool of cancer stem cells that also allows them to form tumours in mice. Irrespective of the model of tumour formation, it is plausible that the individual breast tumour subtypes reflect different cells of origin within the mammary epithelium. The fact that the MaSC-enriched myoepithelial subcompartment lacks ER, PR and HER2 expression led to the speculation that it might also give rise to basal-like breast cancers. In keeping with this notion, BRCA1 mutation carriers often develop basal-like breast cancers, and BRCA1 has been shown to play a role in the regulation of stem cell fate (Foulkes, 2004; Liu et al., 2008). However, recent expression array studies did not confirm the relationship between MaSC and basal breast cancers but demonstrated that the basal BRCA1-associated cancer subtype showed a marked similarity with the luminal progenitor subpopulation (Herschkowitz et al., 2007; Lim et al., 2010). This is consistent with the recent finding by Molyneux et al., who showed that targeted deletion of Brca1 in mouse mammary epithelial luminal progenitors produced tumours that resemble human BRCA1-associated breast cancers with a basal phenotype, while Brca1 deficiency in the myoepithelial compartment led to the formation of tumours that do not phenotypically resemble BRCA1-associated basal human breast tumours (Molyneux et al., 2010). This is at odds with the previous observation that Bera1 mutant tumours originate in the basal subcompartment (Liu et al., 2007). This discrepancy may be explained by differences in the mouse model and method of targeted deletion used. Support for Molyneux’s finding comes from the Visvader group, who demonstrated that human breast tissue heterozygous for a BRCA1 mutation contains an increased luminal progenitor population. They further showed that tissue from BRCA1 mutant carriers as well as tissue from basal-like breast cancers were more similar in their gene expression profiles to the luminal progenitor subset than to any other epithelial subpopulation (Lim et al., 2009). The answer to the question of the cell of origin of luminal breast cancers appears more straightforward as the expression signature of differentiated normal luminal cells exhibits great similarity with that of luminal A and B subtypes (Lim et al., 2010), suggesting that luminal tumours arise from epithelial cells that are committed to a luminal cell fate, presumably at any stage of differentiation. However, based on available data it cannot be ruled out that ER+ tumours may also arise from an ER- stem cell population whose progeny differentiate into ER+ tumour cells with a luminal...
phenotype. As mentioned above, a fraction of HER2-positive tumours can be reclassified as molecular apocrine, thus being defined by the presence of androgen signalling rather than their HER2 status. Irrespective of the exact classification though, current evidence suggests that both molecular apocrine and HER2-positive breast tumours arise from epithelial cells committed to the luminal lineage. Lastly, Perou and colleagues have reported that the MaSC signature has the greatest overlap with the claudin-low breast cancer subtype, which is characterised by the expression of EMT, immune response and stem cell genes, suggesting a basal origin for this subtype (Herschkowitz et al., 2007; Prat et al., 2010). Perou and co-workers have extensively discussed and characterised the claudin-low subtype but few other groups have reported it in human tumour samples. It appears to be a subset of basal-like or normal-like tumours that is most easily seen in cell lines.

A model of the current understanding of the cell of breast cancer origin is depicted in Figure 4.

Figure 4. The cell of origin of the different breast cancer subtypes.
Schematic model of the epithelial hierarchy in the normal human breast and putative relationships with breast cancer subtypes. The different subtypes are grouped with their closest normal epithelial counterpart based on gene expression analysis.
1.5 **A model system for human breast cancer**

The lack of good experimental models that faithfully recapitulate the normal human mammary epithelial lineages and the different breast cancer subtypes is a major barrier to understanding breast cancer and developing better treatments. Commonly used model systems include established breast carcinoma cell lines, xenograft models and genetically engineered mice, all of which have their strengths and weaknesses. While established human breast cancer cell lines have been widely used to study the molecular biology of breast cancer, it is controversial how faithfully they represent primary tumours as they often derive from advanced-stage tumours and pleural effusions. Another weakness is that long-term propagation in culture is likely to have caused them to adapt to culture conditions and to accumulate further mutations and genomic changes over time (Hampton et al., 2009). These limitations also make them a poor model for the study of the normal human breast. Primary human epithelial cells are the ultimate model system to investigate the biology of the normal human breast, and defined transformation of normal human cells should produce tumour cells that resemble human primary tumours more closely than established cell lines that have been cultured over decades.

Transgenic mouse models faithfully recapitulate a wide range of properties of the human disease, as gene expression and phenotypic analyses demonstrate (Lim et al., 2010). However, despite the conservation between the two species, important differences exist. For instance, mouse luminal progenitor cells express significantly lower levels of ERα than their human counterparts (Visvader, 2009), and endogenous serum estradiol levels in sexually mature female mice are comparable to those found in post-menopausal women (Anderson et al., 1998; Laidlaw et al., 1995). These differences in estrogen signalling could potentially reduce the relevance of mouse models to human luminal breast cancer.

Human-in-mouse xenograft models involving the injection of dissociated human mammary cells or the transplantation of fractions of mammary tissue (organoids) into the fat pad of immunocompromised mice allow for the study of human cells in the context of a whole organism. Furthermore, orthotopic xenograft assays are considered the gold standard to assess the tumourigenic, or cancer stem cell, potential of a given cell.
Taking into consideration the advantages and limitations of the different currently available model systems to study the biology of the normal breast and breast cancer, it is conceivable that models derived from normal human primary cells will reproduce the features of the normal breast and breast cancer most faithfully. Such model systems are discussed in detail in the following section.

1.5.1 Human mammary epithelial cell culture

The difficulties associated with the propagation of human mammary epithelial cells in culture have hampered the creation of a faithful model system for human breast cancer. Protocols vary between laboratories but normal human primary mammary epithelial cells are generally obtained from reduction mammoplasty tissue, which, after gross macroscopic examination by a pathologist, is mechanically dissociated and enzymatically digested to remove the extracellular matrix. The resulting small fragments of epithelial tissue, termed organoids, are separated from stromal components of the mixture (including fibroblasts and adipocytes) by gentle centrifugation. Organoids can then be plated directly onto tissue culture plastic or alternatively epithelial fragments can be digested enzymatically further in order to obtain dissociated single cells. While standard media are convenient for the culture of established tumour cell lines, they are not chemically defined due to the supplementation with fetal bovine serum as a source of growth factors, anti-oxidants and vitamins. Human mammary epithelial cells quickly undergo growth arrest after only a few passages when maintained in serum-containing medium unless plated on a layer of fibroblast feeder cells (Smith et al., 1981). Stampfer and colleagues were the first to successfully culture human mammary epithelial cells (HMEC) in chemically defined medium (MCDB 170) for several passages by directly plating organoids and then subculturing single HMEC growing out from adherent tissue fragments (Hammond et al., 1984). However, HMEC cultured in this way still undergo growth arrest after about 5 to 20 populations doublings. This state of “stasis”, commonly known as “culture shock”, is associated with a senescent phenotype and the expression of p16 (CDKN2A; p16INK4A), which is thought to be induced by perturbed growth factor signalling and oxidative stress due to suboptimal culture conditions (Hammond et al., 1984; Romanov et al., 2001; Yaswen and Stampfer, 2001; Yaswen
and Stampfer, 2002). When cultured in serum-free medium, a small variant subset of HMEC (vHMEC) is able to spontaneously overcome senescence by silencing the p16^{INK4A} promoter through methylation (Holst et al., 2003). Variations in p16 expression are also observed in breast tissue in situ, and it has been proposed that silencing of the promoter is an early step during malignant transformation (Tlsty et al., 2004). Despite the suppression of p16, vHMEC invariably stop proliferating, triggered by telomere shortening (Stampfer et al., 1997). It is thus of immediate interest for the study of normal as well as transformed primary mammary epithelial cells to develop cell culture conditions that allow for long-term propagation in vitro without inducing a senescence. Since the pioneering experiments carried out by Stampfer and colleagues using chemically defined medium, it has become evident that the composition of a medium has not only profound effects on the lifespan of primary mammary cells but also determines which epithelial cell type predominates after repeated passaging. Generally, it appears that conventional serum-free media favour the proliferation of myoepithelial cells, while the percentage of luminal epithelial cells already decreases after a single passage (Petersen and van Deurs, 1988; Stingl et al., 1998; Taylor-Papadimitriou et al., 1989). Dontu and co-workers showed that dissociated mammary epithelial cells give rise to so-called mammospheres when cultured in suspension in defined medium that had originally been used to maintain neural stem cells for multiple passages (MEGM, supplemented with EGF, bFGF, heparin and the serum-free B27 mixture of growth factors and vitamins) (Brewer et al., 1993; Dontu et al., 2003). As mentioned above, mammospheres are enriched for early progenitors and possibly MaSC, which are able to differentiate into cells of the different epithelial lineages when subsequently cultured under adherent conditions. Although nonadherent conditions allow for the propagation of progenitor cells, the cells' capacity to self-renew and proliferate is exhausted after approximately four passages. Nevertheless, many laboratories have used MEGM as the conventional medium to propagate HMEC. Our group has obtained essentially the same results using a very similar medium for mammosphere culture that is based on Hepes-buffered Dulbecco’s Modified Eagle Medium (DMEM)/Ham F-12, supplemented with EGF, bFGF, heparin and B27 (Duss et al., 2007). In addition, we routinely propagate transformed HMEC in a modified version of the mammosphere medium additionally containing insulin, o-phosphoethanolamine, apotransferrin, hydrocortisone and isoproterenol (HMM+.
medium). Notably, MEGM and HMM+ media may allow for the propagation of transformed HMEC for several population doublings, but longer-term culture in these media compromises the phenotype of HMEC. Evidence for this cell culture-related problem comes from the observation that tumours derived from xenograft assays of transformed HMEC commonly are of squamous differentiation, a phenotype rarely seen in human breast cancer (Elenbaas et al., 2001). A recent study from the Weinberg lab provides further evidence for the notion that adaptation to the *in vitro* culture conditions is a major determinant of the tumour phenotype observed in xenografts later on (Ince et al., 2007). By culturing primary human mammary epithelial cells (termed breast primary epithelial cells, BPEC, to distinguish them from HMEC grown in conventional medium) in a novel defined medium (so-called WIT medium) on positively charged plastic (Primaria, BD Biosciences), Ince and colleagues were able to change the phenotype of the derived tumours from squamous to more adenocarcinoma-like, which more faithfully reproduces the predominant phenotype of human primary breast tumours. These findings suggest further that adaptation to the cell culture conditions goes along with a change in cell fate, more specifically a change from a glandular phenotype to a skin-like phenotype. In addition, tumours derived from transformed BPEC cultured in WIT/Primaria conditions are characterised by an abundant stromal response and form metastases in immunodeficient mice (McAllister et al., 2008). Importantly, the total culture time and number of passages prior to xenografting of transformed HMEC into mice appears to influence their metastatic potential (our unpublished observations), which is consistent with the importance of cell culture conditions for the cellular phenotype observed by Ince and co-workers. A second medium that is currently widely used is EpiCult-B medium, a commercially available medium for the culture of bipotent progenitor, myoepithelial and luminal cells (Stingl et al., 2006). Unfortunately, Stemcell Technologies, the supplier of EpiCult-B medium and MammoCult medium, which has been optimised for the culture of mammospheres and tumourspheres, has not disclosed the ingredients, and data on long-term culture of primary cells in these media are not available.

Considering the various culture media and techniques, it is evident that when working with human primary mammary epithelial cells, at present the most reliable way to avoid cell culture artefacts is still to minimise the *in vitro* culture time.
1.5.2 Modelling human breast cancer by quantitative transformation of normal human mammary epithelial cells

As discussed above, the presence or absence of expression of the estrogen receptor $\alpha$ is a central characteristic of breast tumours, as it determines not only the pathology of a tumour but also the choice of treatment. Thus, the availability of adequate models of luminal breast cancer is crucial to study the biology of this subtype of the disease and ultimately to optimise therapies targeting ER$\alpha$-dependent signalling pathways. However, to date, models of ER$\alpha$-positive breast cancer are limited to established breast cancer lines, such as the widely used MCF-7 cell line, which produce tumours in mice that, at least in part, resemble the human disease but which also bear the aforementioned caveats. Most transgenic mouse models are ER$\alpha$-negative, but some, such as MMTV-PyMT and the C3(1)/SV40 T-antigen transgenic mouse do produce ER$\alpha$-positive tumours (Fantozzi and Christofori, 2006; Vargo-Gogola and Rosen, 2007). However, tumourigenesis in both models is based on tumour virus-dependent transformation. Thus, available ER$\alpha$-positive human breast cancer lines and transgenic models bear the limitation that they are genetically undefined, which means that the exact causes of tumour formation are difficult to determine.

Problems associated with long-term in vitro culture aside, oncogenic transformation of normal human mammary epithelial cells (HMEC or BPEC) and subsequent xenografting in immunocompromised mice represent an attractive alternative to the aforementioned model systems. The clear advantage of transformed primary human mammary cells is that cells genetically resemble their normal human counterparts as closely as possible if maintained in culture for a short time. At the same time, the expression of a defined set of oncogenes allows for the “quantitative transformation” of normal cells, which means that the effect of oncogene expression is directly reflected in the resulting tumour phenotypes. Previous studies have shown that viral-based expression of a combination of genes, including simian virus 40 small- and large-T antigen (SV40 S/LT), constitutively active RAS, the stabilised MYC T58A mutant, a dominant-negative form of p53 and the telomerase catalytic subunit (TERT) results in the oncogenic transformation of HMEC and tumour formation in mice (Elenbaas et al., 2001; Hahn et al., 1999; Kendall et al., 2005; Rangarajan et al., 2004). However, the transformed HMEC and resulting tumours were ER$\alpha$-negative as ER$\alpha$ is rapidly lost in culture. Our group has previously created an ER$\alpha$-positive
tumour model by defined transformation of normal HMEC by expression of ESR1, TERT, BMI1 and MYC (Duss et al., 2007). The transformed HMEC produced ERα-positive tumours that were estrogen-dependent for growth, sensitive to the ERα antagonist fulvestrant, and metastasised to multiple organs, thus reproducing several characteristics of human luminal breast cancer. Forced expression of ERα was ensured by lentiviral-based expression of the gene, and ERα-dependent growth arrest and terminal differentiation of HMEC was prevented by expression of the polycomb gene BMI1. BMI1 was chosen as it is overexpressed in luminal primary breast tumours and because of its role in stem cell self-renewal and repression of p53- and retinoblastoma protein (RB)-induced senescence by silencing the CDKN2A locus. BMI1 further inhibits MYC-induced apoptosis through the same targets (Jacobs et al., 1999b). The tumours that formed in the mouse mammary fat pad were a mixture of invasive ductal carcinoma and squamous carcinoma. Squamous differentiation is rare in human breast cancer but common in HMEC xenograft assays (Elenbaas et al., 2001; Ince et al., 2007). As discussed above, this change in cell fate is most likely a consequence of adaptation to a suboptimal in vitro culture environment. Ince and colleagues thus proposed that human mammary epithelial cells maintained in WIT/Primaria conditions and transformed with TERT, SV40 large-T antigen and oncogenic RAS form adenocarcinomas in their xenograft model, while cells cultured in conventional conditions give rise to tumours of predominantly squamous differentiation (Ince et al., 2007). However, the authors gave no explanation of what exactly may cause WIT/Primaria-derived BPEC to maintain their original glandular phenotype. Further evidence for the importance of cell culture conditions comes from the observation that transformed HMEC increasingly lose their metastatic potential after long-term in vitro culture (Stephan Duss, unpublished observation). Alternatively, the discrepancy in tumour phenotype may be due to differences between the microenvironment of the human breast and mouse mammary gland (Parmar and Cunha, 2004). “Humanising” the mammary gland by co-injection of human mammary fibroblasts (HMF) with transformed HMEC did not change the observed tumour phenotype (Duss et al., 2007). However, normal HMF may not accurately resemble the paracrine environment of tumour-associated stroma in primary human tumours (Karnoub et al., 2007; Orimo et al., 2005). Related to the question of the tumour cell environment is the choice of xenograft model. While adult
mammary stem cell assays generally require the “clearing” of the mouse mammary fat pad, HMEC-derived tumour cells can be xenografted directly by injection into the forth inguinal mammary gland (Duss et al., 2007). An alternative to orthotopic xenografts is injection under the renal capsule or subcutaneously. A recent comparative xenograft study suggests that the site of injection (subcutaneous versus third or forth mammary gland) impacts the size and vascularisation of the resulting tumours and that optimal tumour take occurs with orthotopic injections into the forth gland (Fleming et al., 2010). In an effort to mimic non-invasive human ductal carcinoma in situ (DCIS), Medina and colleagues recently developed an intraductal human-in-mouse xenograft model (Behbod et al., 2009). This is an intriguing model system as it reproduces the physiological site of ductal tumour origin far more closely than injections into the stromal part of the mouse mammary gland. The original study only reported the successful engraftment of malignant and DCIS cell lines and it remains to be determined if this potentially very valuable model system supports the growth of transformed primary cells.

Another possible explanation for the observed phenotypic differences between human luminal breast cancers and the tumours in our ERα-positive mouse model may be that ERα expression alone is not sufficient to define the luminal subtype. As mentioned above, gene expression analysis of primary breast tumours shows the co-expression of a cluster of genes that are consistently co-expressed with ERα in luminal A tumours, including GATA3, FOXA1 and XBP1 (Sorlie et al., 2001). Thus, it is likely that ERα requires co-factors other than BMI1 and MYC to activate the full range of its target genes in luminal tumours. Transcriptional regulators are therefore of particular interest as candidate genes that regulate luminal cell fate, both in the normal mammary gland and in luminal tumours, and a combination of our model with the improved culture conditions described by the Weinberg lab is potentially an excellent model system to study the oncogenic potential of such genes of interest.

1.6 Identification of new mammary oncogenes and tumour suppressor genes

Advances in technology now allow for high-throughput analyses to identify new oncogenes and tumour suppressors but the increasing sophistication of these
techniques also impedes the interpretation of the resulting complex data. Sequencing studies, first of candidate genes, then of classes of genes, then of all exons and finally of whole genomes have identified cancer-associated point mutations in genes such as BRAF and ERBB2 (reviewed by Stratton et al., 2009) but the vast number of mutations discovered by recent whole cancer genome sequencing efforts (Forbes et al., 2010) is itself a major challenge because it is difficult to distinguish causative oncogenic “driver” mutations from “passenger” mutations.

A second technique that has been employed to study genetic alterations associated with breast cancer is the analysis of copy number changes. Comparative genome hybridisation (CGH) analysis and similar genomic array methods have been used successfully to identify copy number alterations in known breast cancer genes such as CCND1, MYC, ERBB2, TBX2 as well as previously unknown genes located in amplicons (Chin et al., 2006; Chin et al., 2007; Hicks et al., 2006).

Human genetics represents an alternative and possibly more reliable way to identify new breast cancer genes as it directly associates breast cancer risk with the presence of a genomic variation. Classic linkage studies that identified breast cancer susceptibility genes such as breast cancer susceptibility genes 1 and 2 (BRCA1 and 2) (Hall et al., 1990; Wooster et al., 1994) have been succeeded by large-scale genome-wide association studies (GWAS) in recent years (Easton et al., 2007; Stacey et al., 2007; Stacey et al., 2008; Walsh and King, 2007). Given a sufficient number of samples, GWAS are a powerful tool for the discovery of genetic variants that are associated with an increased breast cancer risk but which would go otherwise unnoticed due to their low penetrance. The discovery of a large number of new low-penetrance breast cancer susceptibility genes may not bear immediate importance for clinical prevention at this point. However the identification of previously unknown susceptibility genes is valuable for the understanding of breast oncogenesis as a positive family history is still considered the most important risk factor for the development of the disease, underscoring the importance of hereditary factors. Mutations in the known high-penetrance breast cancer susceptibility genes including BRCA1, BRCA2, TP53 and PTEN account for less than 25% of familial breast cancer cases, suggesting the existence of a number of low-penetrance genes that have yet to be identified (Easton, 1999; Walsh and King, 2007). In recent years, several DNA damage repair and cell cycle checkpoint genes such as CHEK2 and PALB2 have
emerged as such low-penetrance factors (Mavaddat et al., 2010a; Walsh and King, 2007).
In the search for new low-risk breast cancer genes, several independent GWAS have
recently identified five loci containing single nucleotide polymorphisms (SNPs) that
are strongly associated with increased breast cancer risk (Easton et al., 2007; Hunter
et al., 2007; Stacey et al., 2007). Four of these loci contain plausible causative genes:
fibroblast growth factor receptor 2 (FGFR2), mitogen activated protein 3 kinase 1
(MAP3K1), lymphocyte-specific protein 1 (LSP1) and TOX high mobility group box
family member 3 (TOX3). Experimental evidence confirming a role in breast
tumourigenesis has been most convincing for the receptor tyrosine kinase FGFR2 thus
far. The FGFR2 gene is frequently amplified in breast tumours (Adelaide et al., 2007;
Adnane et al., 1991). FGFR2 is able to transform normal mammary epithelial cells
(Moffa et al., 2004), while breast cancer cell lines bearing an FGFR2 amplification
are highly sensitive to FGFR2 inhibitors (Turner et al., 2010). Follow-up studies have
confirmed the association of common variants in the FGFR2 gene with increased
breast cancer risk, in particular with ER-positive cancer (Meyer et al., 2008; Reeves et
al., 2010; Stacey et al., 2008), possibly due to increased FGFR2 expression caused by
a mutation in the binding site for a transcriptional activator in the FGFR2 promoter
(Meyer et al., 2008). A recent study by Rosen and colleagues demonstrated that
FGFR signalling accelerates tumour formation in a MMTV-Wnt1 mouse model (Pond
et al., 2010), confirming further that GWAS are useful for the identification of
biologically relevant breast cancer genes.

1.7 Thymocyte selection-associated high mobility group box protein
family member 3 (TOX3)

1.7.1 Identification of TOX3 as a breast cancer susceptibility gene in GWAS

The second new candidate breast susceptibility gene of great interest is TOX3. The
synonymous SNP rs3803662 was strongly associated with increased breast cancer
risk in the original GWAS (Easton et al., 2007; Stacey et al., 2007). It is located on
chromosome 16q12.1, 8 kb upstream of the TOX3 coding region. At the same time it
lies in the fourth exon of the predicted gene LOC643714 (Fig. 5). It cannot be ruled
out that the uncharacterised hypothetical gene LOC643714 is the causative gene that is associated with increased breast cancer risk (Ruiz-Narvaez et al., 2010). However, we found TOX3 (formerly called trinucleotide containing 9, TNRC9) to be the more convincing candidate gene as it is expressed in the gene cluster that defines luminal tumours (see Results), which is consistent with the findings of Stacey and colleagues, who showed in their original GWAS that SNP rs3803662 was exclusively associated with an increased risk of ER-positive breast cancer (Stacey et al., 2007). A follow-up study also confirmed the upregulation of TOX3 expression in luminal tumours (Nordgard et al., 2007). At the time of the publication of the original GWAS, TOX3 gene function was uncharacterised although sequence predictions had previously placed it in the TOX family of HMG-box transcription factors, as discussed below (O'Flaherty and Kaye, 2003). TOX3 mRNA upregulation had been shown in response to expression of severe acute respiratory virus syndrome corona virus (SARS-CoV) 3a protein in lung epithelial cells but the significance of this finding is unclear (Tan et al., 2005). TOX3 had also previously been identified as a gene potentially involved in metastasis of breast cancer to the bone, a characteristic of ER-positive breast tumours (Smid et al., 2006), further suggesting a role for TOX3 in luminal breast cancer.

![Figure 5](image)

Figure 5. The TOX3 locus.

Linkage disequilibrium plot of the 16q12 locus which encompasses the TOX3 and LOC643714 genes. The arrow marks SNP rs3803662. Figure taken from (Udler et al., 2010).

Mounting evidence has since implicated TOX3 in breast cancer. A multitude of follow-up studies using large sample numbers have investigated the correlation
between the common variants found in the five novel breast cancer susceptibility loci and clinical parameters including tumour grade, steroid receptor status, node involvement, disease onset, survival and family history of breast cancer (Garcia-Closas et al., 2008; Latif et al., 2009; Long et al., 2010; Mavaddat et al., 2010b; Tapper et al., 2008; Udler et al., 2010). A recent meta-analysis of previous studies taking into account 62,000 subjects confirmed a significant correlation between SNP rs3803662 and increased breast cancer risk (Chen et al., 2011). The results consistently confirm a role for TOX3 in breast oncogenesis. Fine-scale mapping of the 16q12 locus has identified additional SNPs in this region that are associated with increased breast cancer risk, albeit less strongly than rs3803662, underlining the significance of earlier findings. The majority of variants identified by Udler and colleagues in the latter study were located in the intergenic region between the TOX3 and LOC643714 (Fig. 5). They further showed that this intergenic region is highly conserved and, based on DNase I hypersensitivity profiles, has an open chromatin conformation consistent with active transcription (Udler et al., 2010). It should be noted that studies addressing the association of TOX3 with ER status (specifically an increased risk for ER-positive breast cancer) are somewhat contradictory, although the majority of follow-up studies has thus far confirmed the correlation (Garcia-Closas et al., 2008; Mavaddat et al., 2010b; Tapper et al., 2008).

As mentioned above, the TOX3 gene is located on the long arm of chromosome 16. Notably, loss of heterozygosity (LOH) on 16q12 is one of the most frequent genetic events in breast cancer, suggesting the presence of one or more tumour suppressor genes in this area (Rakha et al., 2006). SNP rs3803662 was shown to correlate with disease onset before the age of 60 years (Huijts et al., 2007). It is thus tempting to speculate that TOX3 may be one of the tumour suppressor genes on 16q12 and that the presence of the SNP causes an increase in breast cancer risk due to downregulation of TOX3 expression.

Taken together, the existing data implicate TOX3 in the development of ER-positive breast cancer, making it an interesting candidate gene to test in our luminal tumour model.
1.7.2 The high mobility group (HMG) box protein family

Based on sequence alignment, TOX3 belongs to the TOX subfamily of high mobility group (HMG) proteins (O'Flaherty and Kaye, 2003). The HMG protein family is a diverse superfamily of nonhistone chromosomal proteins that were discovered in mammalian cells more than 30 years ago. The HMG proteins were originally named based on their unusually rapid gel electrophoretic mobility compared to other chromatin proteins. This is caused by their high content of charged amino acid residues, which also explains their extractability from chromatin in 350 mM NaCl and 5% perchloric- or trichloroacetic acid (Goodwin and Johns, 1973; Goodwin et al., 1973; Reeves, 2010). The recently reclassified ‘HMG-motif proteins’ are subdivided into three superfamilies based on the sequence of their DNA binding domain: the functional motif of the HMGB family is the ‘HMG-box’, that of the HMGN family is called ‘nucleosomal binding domain’ and that of the HMGA family is termed ‘AT-hook’ (Bustin, 2001). While each of the different subfamilies has a unique protein signature and a characteristic functional DNA binding sequence motif, all HMG proteins have in common that they bind to specific structures in DNA or chromatin with little if any specificity for the target DNA sequence (Bustin, 1999). The functional DNA binding motifs that are characteristic of each HMG subfamily have been identified in numerous other nuclear proteins that interact with chromatin but these differ from the canonical HMG family members in several aspects. The archetypal HMG proteins are ubiquitous and generally expressed in all cells of higher eukaryotes; they are relatively abundant and bind to DNA in a sequence-independent manner. In contrast, non-canonical HMG proteins are cell type-specific, not abundant and often bind to DNA with sequence specificity. In addition to the HMG motif, the latter proteins frequently contain other non-HMG functional motifs (Bustin, 1999).

Based on sequence alignment, TOX3 best fits into the HMGB (or HMG-box) family (O'Flaherty and Kaye, 2003). I will thus focus my discussion on this HMG subfamily. As mentioned above, all HMGB proteins possess a DNA-binding motif related to the HMG-box that was originally identified in HMGB1 and HMGB2 (previously termed HMG1/2) (Goodwin and Johns, 1973). Beyond the similarity in their DNA-binding motifs, the HMG-box family members are classified into either of two subgroups distinguished by their abundance, function, DNA specificity and number of HMG-boxes. In humans, the first subgroup contains proteins that comprise two HMG-boxes
(HMGB1-4, discussed below) and the RNA polymerase I transcription factor UBF with four to six HMG-boxes (Stros et al., 2007). Proteins in this subgroup generally have two or more HMG-boxes, are abundant in all cell types and bind to DNA and chromatin without sequence specificity. The second subgroup consists of a multitude of highly diverse proteins that predominantly contain a single HMG-box, are cell type-specific and much less abundant. They recognize specific DNA sequences, albeit with relatively low specificity due to the limited number of hydrogen bonds that can be formed within the minor groove (Bustin, 1999; Stros, 2010; Stros et al., 2007). This subgroup includes family of the T cell transcription factors/lymphoid enhancer-binding factors (TCF/LEF), sex-determining factors SRY and related SOX proteins, chromatin remodelling factors SMARCE1/BAF57 and PB1, as well as the T lymphocyte differentiation factor TOX (reviewed in Stros et al., 2007).

The human HMGB family has three canonical members, HMGB1, HMGB2 and HMGB3, and the recently identified HMGB4 (Catena et al., 2009), which are encoded by individual genes, are abundant in most cell types and bind to DNA without sequence specificity. HMGB1-3 are approximately 25 kDa in molecular mass, share 80% sequence identity, and are highly conserved across species. All four proteins contain a bipartite DNA binding motif encompassing an N-terminal HMG-box A and a central HMG-box B (Fig. 6 A). The HMG-box consists of approximately 80 amino acids that form a characteristic, twisted, L-shaped fold formed by three α helices, which binds to the minor groove of AT-rich B-DNA, thereby bending the DNA (Bustin, 1999 and references therein) (Fig. 6 B). HMG-box B, more than HMG-box A, is predominantly responsible for DNA bending in HMGB1/2 proteins (Paull et al., 1993; Teo et al., 1995). A second characteristic of the canonical HMGB1-3 proteins (but which is not present in HMGB4) is their acidic tail (Fig. 6 A). HMGB1 has been shown to interact with a multitude of proteins including transcription factors, viral proteins and DNA repair proteins via its HMG-boxes, while the acidic tail mediates the interaction with histones, e.g. histone H1. The interaction with chromatin and other nuclear proteins via the HMG boxes and concomitant binding to histones via the acidic tail thus enable HMG-box proteins to function as architectural proteins and to promote the formation of complex nucleoprotein structures (Gerlitz et al., 2009). HMGB binding to other proteins has been observed in vitro and in vivo but it has been suggested that the presence of chromatin is required for at least some of these interactions (Stros, 2010). The interaction of HMGB proteins with other factors
is beyond the scope of this chapter but has been extensively reviewed elsewhere (Stros, 2010; Stros et al., 2007 and references therein). However, it is of interest that HMGB1 interacts with all class I steroid receptors (Reeves, 2010). In particular, it interacts with ERα and promotes its binding to the estrogen response element (ERE) in a dose-dependent manner (Verrier et al., 1997; Zhang et al., 1999). Furthermore, it facilitates low-specificity binding of ERα to half-site binding elements to an extent that resembles high affinity consensus binding (Das et al., 2004; El Marzouk et al., 2008).

Figure 6. Model of HMG-box structure and function.

(A) Amino acid sequence HMGB1 (accession CAE48262) and domain structure of the canonical HMGB family member HMGB1. The HMG boxes A and B are indicated by a red and blue box, respectively; the positions of the individual α-helices constituting the HMG boxes are indicated by red and blue bars, respectively. The acidic tail is marked by an orange box. (B) Structure of the HMG box A of HMGB1 (left) (Hardman et al., 1995) and of the HMG box B of SRY (Stott et al., 2006), as determined by NMR spectroscopy. (C) Putative modes of action of HMGB proteins as architectural factors in transcription. The HMGB factor interacts with a transcription factor (TF1, top left) and recruits it to its DNA binding site, which has been pre-bent by the HMGB factor, to form an intermediate ternary complex (bottom left). Another factor (TF2) is recruited to the TF1-DNA-HMGB complex and the HMGB factor is released (bottom right). Alternatively, the DNA is bent by HMGB resulting in enhanced binding of the specific transcription factors without direct interaction with the HMGB factor (top right). Figures taken from (Stros, 2010).
In keeping with the number of their interacting proteins and chromatin-binding properties, a multitude of functions has been ascribed to HMGB proteins, albeit important differences exist between the abundant HMGB1 and its closely related proteins and the cell type-specific HMGB factors. HMGB1 in particular has been implicated in various biologically important processes including transcription, chromatin remodelling including DNA repair, DNA replication, translocation between nucleus and cytoplasm as well as V(D)J recombination. In addition to its nuclear function, HMGB1 can also act as a cytokine and plays a role in necrosis, inflammation and apoptosis (Scaffidi et al., 2002). Considering its various functions, it is not surprising that HMGB1 is the most abundant nonhistone protein found in the nucleus with approximately 1 molecule per 10-15 nucleosomes (Reeves, 2010). Its abundance also allows a fraction of HMGB1 protein to always be in contact with chromatin despite its high mobility within the nucleus. There are several models of how nuclear HMGB proteins interact with chromatin and protein partners. The HMG-box is a versatile DNA binding motif that can bind to already distorted DNA structures with high affinity including four-way junctions, bulges, kinks and modified DNA such as cisplatin adducts (Pil and Lippard, 1992), which is consistent with a function in the recognition and repair of damaged DNA. It can also bind to and then bend DNA, thus making the chromatin more accessible for transcription factors and other chromatin-remodelling proteins. It is unclear at this point whether HMGB proteins facilitate access to DNA by forming intermediate tertiary complexes with the DNA and other transcription factors or by first distorting the DNA, thereby allowing subsequent binding of other proteins (Fig. 6 C).

1.7.3 The thymocyte selection-associated HMG box (TOX) protein family

As discussed above, in addition to the canonical HMGB proteins, the HMGB family comprises a plethora of non-canonical, cell type-specific nuclear proteins, including the thymocyte selection-associated HMG-box (TOX) protein family named after the thymocyte transcription factor TOX. Based on sequence predictions, the TOX family contains four proteins, TOX1-4, as well as two pseudogenes in the mouse. The members of the TOX family are approximately 500-600 amino acids and 60-70 kDa in size. TOX3 in particular is predicted to encode a 576-amino acid protein with a
calculated molecular mass of 63.3 kDa (O'Flaherty and Kaye, 2003). The founding member of this HMGB subfamily is the murine Tox gene, which was identified by Kaye and co-workers in a gene expression study investigating thymocyte differentiation (Wilkinson et al., 2002). The function of murine Tox is discussed in more detail below. Human TOX had previously been cloned from brain but its function remains to be characterised in human cells (Nagase et al., 1998).

The family member TOX2 (also called GCX-1) was cloned from a rat ovarian granulosa cell cDNA library. It is exclusively expressed in reproductive tissues involving the hypothalamic-pituitary-gonadal axis and appears to function as a transcriptional activator during follicular development (Kajitani et al., 2004).

Aside from its identification as a putative breast cancer predisposition gene in GWAS, TOX3 was recently identified and cloned from a rat cortical cDNA library in a screen designed to find new transcriptional transactivators. The authors showed that TOX3 interacts with both cAMP response element (CRE)-binding protein (CREB) and CREB-binding protein (CBP) to activate transcription in neurons in a calcium-dependent manner (Yuan et al., 2009). Recently, TOX3 was further identified as a downstream target of the orphan G-protein-coupled-receptor GPR39, whose overexpression promotes neuronal survival by induction of serum-response element (SRE)-containing promoters. Along with TOX3, the transcriptional co-activator CITED1 was induced by GPR39 (Dittmer et al., 2008). Consistent with a role as a survival factor, TOX3 activated the expression of pro-apoptotic genes and downregulated anti-apoptotic genes. The study further showed that TOX3 interacted with phosphorylated CREB to activate CRE-containing promoters and with CITED1 to activate ERE-containing promoters (Dittmer et al., 2010). The implication of these findings for TOX3 function in the breast is addressed below.

The fourth family member, TOX4 (also termed LCP1), was identified in a yeast two-hybrid screen as an interacting partner of protein phosphatase 1, regulatory (inhibitor) subunit 10 (PPP1R10, also called PNUTS). Using co-immunoprecipitation and TOX4 deletion constructs, TOX4 was shown to interact with PPP1R10 through its extreme C-terminus, and it co-localises with PPP1R10 in HeLa cells. It was further shown to act as transcriptional activator in HEK 293T cells. A second study confirmed the interaction between TOX4 and PNUTS, which form a stable multimeric complex with protein phosphatase 1 (PP1) and WDR82, a regulatory subunit of the SETD1 Histone histone 3 lysine 4 methyltransferase complex in HEK293T cells. This finding
confirms a potential role for TOX4 in the regulation of transcriptional activation (Lee et al., 2010). Notably, the region of TOX4 that is required for binding to PNUTS is not conserved between TOX4 and TOX3 (Lee et al., 2009). Furthermore, a recent screen for proteins binding to damaged DNA has shown that a complex containing TOX4 and PNUTS is able to recognise DNA adducts caused by platinating anticancer agents (Puch et al., 2010), a known property of other members of the HMG box family. The currently known functions of the TOX family members are summarised in Table 2.

Table 2. Human TOX family members and their murine homologues.

<table>
<thead>
<tr>
<th>Human gene (official gene symbol)</th>
<th>Human gene aliases</th>
<th>Chromosome location</th>
<th>Murine homologue</th>
<th>Identification, biological functions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOX</td>
<td>KIAA0808, TOX1</td>
<td>8q12.1</td>
<td>Tox</td>
<td>Regulator of lineage commitment during T lymphocyte, natural killer (NK) cell and lymphoid tissue-inducer cell development in mice. Regulator of differentiation of human NK cells. Cloned from human brain.</td>
<td>(Aliahmad and Kaye, 2008; Aliahmad et al., 2004; Nagase et al., 1998; O’Flaherty and Kaye, 2003; Wilkinson et al., 2002) (Aliahmad et al., 2010; Yun et al.)</td>
</tr>
<tr>
<td>TOX2</td>
<td>GCX-1</td>
<td>20q13.12</td>
<td>Tox2</td>
<td>First cloned from rat ovarian granulosa cell cDNA library. Transcriptional activator, exclusively expressed in reproductive tissues.</td>
<td>(Kajitani et al., 2004)</td>
</tr>
<tr>
<td>TOX3</td>
<td>TNRC9, CAGF9</td>
<td>16q12.1</td>
<td>Tox3</td>
<td>Identified as breast cancer predisposition gene in GWAS. First cloned from rat cortical cDNA library. Regulation of calcium-dependent transcription via interaction with CREB and CBP in neurons. Interacts with phosphorylated CREB and CITED1 in neurons to activate transcription of CRE- and ERE-containing promoters.</td>
<td>(Dittmer et al., 2010; Dittmer et al., 2008; Udler et al., 2010; Yuan et al., 2009)</td>
</tr>
<tr>
<td>TOX4</td>
<td>KIAA0737, LCP1</td>
<td>14q11.2</td>
<td>Tox4</td>
<td>Identified in yeast two-hybrid screen as protein phosphatase 1 (PP1) nuclear targeting subunit (PNUTS); transcriptional activator, forms stable multimeric complex with PNUTS, PP1 catalytic subunit and WDR82, disruption of which results in cell cycle arrest and apoptosis. Binds to DNA adducts.</td>
<td>(Lee et al., 2010; Lee et al., 2009; Puch et al., 2010)</td>
</tr>
</tbody>
</table>
1.7.4 **TOX family members are evolutionarily conserved and share common functional domains**

TOX family members are highly conserved between humans and mice. Furthermore, conservation of the TOX subfamily, and the TOX HMG-box motif in particular, extends to evolutionarily more distant species, as TOX-like proteins have also been identified in pufferfish and mosquito. Comparison of TOX family members shows that human and murine homologues share an almost identical HMG-box motif (Fig. 7, 8) (O’Flaherty and Kaye, 2003). Interestingly, the HMG-box motif found in all TOX family members is most closely related to the archetypal HMGB motif of HMGB proteins that bind to DNA without sequence specificity and are ubiquitous and abundant, such as HMGB1 (HMG-box A) and UBF/UBTF (Fig. 7 A). This is surprising as all TOX proteins appear to be expressed in a spatially and temporally regulated fashion at only moderate levels (Dittmer et al., 2010; Kajitani et al., 2004; Lee et al., 2009; O’Flaherty and Kaye, 2003; Wilkinson et al., 2002). For instance, Kaye and colleagues showed in their original study that TOX mRNA expression was highest in the thymus, followed by the liver and brain but TOX was not expressed in other tissues (Wilkinson et al., 2002). In this respect, the TOX family members fit better into the subgroup of non-canonical HMGB proteins that have a single HMG-box and bind to DNA with sequence specificity, such as SRY, SOX or LEF1 (Fig. 7 A). Furthermore, a common consensus sequence within HMG-box motifs, GXXW, is replaced with AXXW in TOX family members (Fig. 7 B). This unique combination of conserved properties may suggest a very specialised function for TOX proteins.

In addition to the conservation between TOX family members, TOX3 itself is evolutionarily conserved across species (Fig. 8). Furthermore, fine scale mapping analysis of the 133 kB region on 16q12 that is associated with increased breast cancer risk, and in particular the putative TOX3 promoter region, showed evolutionary conservation (Udler et al., 2010) (Fig. 5)
**A**

<table>
<thead>
<tr>
<th></th>
<th>230</th>
<th>EKRAPDSGKPKTPKKKKKDPNEPOKPVSAAYALFFRDQAIAIKGQNNATGGEVSKIV</th>
<th>289</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOX</td>
<td>236</td>
<td>EKRPSADMGKPKTPKKKKKDPNEPOKPVSAAYALFFRDQAIAIKGQNNATGGEVSKIV</td>
<td>295</td>
</tr>
<tr>
<td>HMGB1</td>
<td>70</td>
<td>RYERKMTVPPGKETKFFDPNAAKRPSAFLFSCSEYRKIKGDHGLSGDVAKKL</td>
<td>129</td>
</tr>
<tr>
<td>UBTF</td>
<td>87</td>
<td>EILDAQEHYNPKKKLNNKHPDFKKPLFPPFEMKRAKYALKDHEMNLLEKRL</td>
<td>146</td>
</tr>
<tr>
<td>SRY</td>
<td>35</td>
<td>LTCTSCSNQYCTESGKSNQVDKRRMPAFIVWSDPOQRKMALENPRRSEISQOL</td>
<td>94</td>
</tr>
<tr>
<td>SOX4</td>
<td>34</td>
<td>PTPGASTTGKADDPSWCTPSGHHKRPSAHPWNGQIERKIIQEQUDHHAERK</td>
<td>93</td>
</tr>
<tr>
<td>LEF1</td>
<td>74</td>
<td>HTDSLHMVQPPHEQKKEQFKRPHFKFLAPMLKEMRANVAVECTKEEAAINQIL</td>
<td>133</td>
</tr>
</tbody>
</table>

**B**

<table>
<thead>
<tr>
<th></th>
<th>236</th>
<th>EKRPSADMGKPKTPKKKKKDPNEPOKPVSAAYALFFRDQAIAIKGQNNATGGEVSKIV</th>
<th>295</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOX</td>
<td>221</td>
<td>EKRPSADMGKPKTPKKKKKDPNEPOKPVSAAYALFFRDQAIAIKGQNNATGGEVSKIV</td>
<td>295</td>
</tr>
<tr>
<td>TOX</td>
<td>230</td>
<td>EKRAPDSGKPKTPKKKKKDPNEPOKPVSAAYALFFRDQAIAIKGQNNATGGEVSKIV</td>
<td>295</td>
</tr>
<tr>
<td>TOX</td>
<td>189</td>
<td>ORTVVEAGKPPKKKDPNEPOKPVSAAYALFFRDQAIAIKGQNNATGGEVSKIV</td>
<td>257</td>
</tr>
<tr>
<td>TOX</td>
<td>296</td>
<td>ASMDLGEQOKYVKKTEAAKKEVLKALAYRSLVSKAAEASAEQTIRSVQOTLASHYKKYELQAEKREYKYTRPRRAKMLFKCSLLPAPASVL</td>
<td>154</td>
</tr>
<tr>
<td>SOX4</td>
<td>94</td>
<td>GKRKKLKDSDKIPFIREAERLRLKMDYDPRPRKKVKSGNSSSSSAAAASKPRGEE</td>
<td>153</td>
</tr>
<tr>
<td>LEF1</td>
<td>134</td>
<td>GRRHLAESREEQAKYELARKEQLMQFAWARPYNQSGTGRP</td>
<td>193</td>
</tr>
</tbody>
</table>

**Introduction**

Comparison of the TOX and TOX3 HMG-box motifs with those of HMGB proteins shows that the TOX family members are more closely related to ubiquitous HMGB proteins that bind DNA without sequence specificity (above line) than sequence- and cell type-specific HMGB proteins (below line). Similarities between both groups and the HMG-box consensus sequence (Pfam accession number pf00505) are highlighted in grey. Residues in red and green distinguish the HMGB-like HMG-box from the non-canonical HMG-box. The commonly found consensus sequence (G/A)XXW is highlighted in yellow. (B) Comparison of the HMG-box motifs and putative nuclear localisation signals (in purple) of the TOX family members. The AXXW consensus sequence is highlighted in yellow. The conserved exon-exon boundary present in all four genes is marked in turquoise.

Figure 7. Conservation of the HMG-box motif.
Figure 8.  TOX3 is evolutionarily conserved across species.

(A) Alignment of TOX3 amino acid sequences from human, mouse, rat, and clawed frog. (B) Phylogenetic tree showing evolutionary conservation of TOX3 amino acid sequences. Alignment analysis was performed with ClustalW version 2.1; colours reflect physicochemical properties of the amino acids; the ASMW consensus sequence within the HMG-box is underlined.
The TOX family members are less conserved outside of the HMG-box but do share common domains (Fig. 7-9). A conserved lysine-rich putative bipartite nuclear localisation signal (NLS) is located N-terminally adjacent to the HMG-box (Fig. 7 B, 9) (O’Flaherty and Kaye, 2003). The wild-type forms of all four TOX family members localise to the nucleus (Kajitani et al., 2004; Lee et al., 2009; Wilkinson et al., 2002; Yuan et al., 2009), while deletion of this basic region prevents nuclear localisation of TOX2 and TOX4 (Kajitani et al., 2004; Lee et al., 2009). All four family members can act as activators of transcription in mammalian cells (Aliahmad and Kaye, 2006; Dittmer et al., 2010; Kajitani et al., 2004; Lee et al., 2009; Yuan et al., 2009). The transactivation domain was mapped to the N-terminal amino acid residues 25-63 of TOX2 and the N-terminal 250 amino acid residues of TOX4, respectively (Kajitani et al., 2004; Lee et al., 2009). The transcriptional activation ability of full-length rat TOX3 fused to the GAL4 DNA-binding domain is reduced by 50% through deletion of the part of the protein that is C-terminal to the HMG-box. Further deletion of the DNA-binding domain has no effect on transcriptional activation in this context. However, the ability to activate transcription is completely abrogated by deletion of the N-terminal part of the protein, further demonstrating the conservation between the functional domains of the TOX family members (Yuan et al., 2009). TOX activity may be regulated by post-translational modifications, as the N-terminus of TOX is subject to phosphorylation but the exact residues are unknown at this point (Aliahmad and Kaye, 2006).

Further functional motifs that have been identified in the TOX family members include a protein interaction domain at the carboxy-terminus of TOX4 that is required for its interaction with PNUTS (Lee et al., 2009). Notably, the domain that mediates calcium responsiveness of TOX3 was also mapped to the C-terminus, suggesting that this region of the protein is also important for its interaction with CREB and CBP (Yuan et al., 2009). TOX3 interaction with CITED1 is mediated by the HMG-box and the N-terminally adjacent part of the protein (Dittmer et al., 2010). A unique feature of human TOX3 is a CAG trinucleotide repeat coding for a C-terminal glutamine stretch, which is not present in any of the other TOX proteins but conserved between the TOX3 gene in different species (Fig. 8). Longer glutamine stretches have been found in a number of transcription factors including the androgen receptor and TATA binding protein, where they appear to activate transcription (Gerber et al., 1994). Notably, the murine HMGB protein SRY also contains a C-terminal glutamine
stretch, which is required for male sex determination (Bowles et al., 1999). The expansion of CAG repeats has been shown to be associated with neurodegenerative disorders in the past, where they are thought to cause conformational changes and protein misfolding (Margolis et al., 1997; Masino and Pastore, 2002; Ross, 2002). The function of the CAG repeat in the carboxy-terminus of TOX3 and whether it plays a role in breast tumourigenesis remains to be determined. Northern blot analysis showed the presence of a 3-kb and a 5-kb TOX transcript in thymocytes expressing a transgenic cDNA-derived TOX, the former of which was more abundant. In addition, Western blot analysis showed two forms of TOX that were ca. 57 kDa and ca. 63 kDa in size (Wilkinson et al., 2002). Similarly, two transcripts of endogenous TOX3 (3 kb and 5 kb) and TOX4 (4 kb and 6 kb) have been detected (O'Flaherty and Kaye, 2003; Yuan et al., 2009). NCBI GenBank contains two entries for human TOX3 (NM 001080430.2 and NM 001146188.1). Transcript variant 2 differs from variant 1 in the 5’ UTR and coding sequence, and the resulting isoform is predicted to have a shorter and distinct N-terminus compared to isoform 1. It remains to be investigated whether the different isoforms found at the mRNA and protein level represent splice isoforms or post-translational modifications.

Figure 9. TOX protein domains.

Schematic drawing of the four TOX family members that have been experimentally analysed and the domains present in the individual proteins. The exact location of the phosphorylation site in TOX is unknown. Lys, lysine; Pro, proline; Ser, serine; Glu, glutamine; Ca, calcium.
In summary, currently available data describe the TOX proteins as transcriptional activators that contain a central DNA-binding domain, which is flanked on the N-terminal side by a basic nuclear localisation signal and an acidic transactivation domain and on the C-terminal side by a protein interaction domain.

1.7.5 **TOX is a regulator of lineage commitment in the innate and adaptive immune system**

The TOX protein family member that has been studied in most detail is TOX itself. Interestingly, TOX plays an important role in the regulation of lineage commitment both in the adaptive and innate immune system. As parallels are often drawn between the haematopoietic and mammary epithelial stem cell hierarchies, it is tempting to speculate that TOX3 may play a role in the regulation of mammary epithelial cell fate that is similar to the function of TOX in haematopoiesis.

Thymocytes are immature T cell precursors that are derived from haematopoietic progenitor cells in the bone marrow, which migrate through the blood stream to the thymus, where they develop into mature T cells. T cell maturation is a complex process of selection and differentiation of functional T cells in an antigen-dependent manner, which involves several rearrangements of the T cell receptor (TCR) gene, resulting in a pool of T cells that are able to respond to foreign pathogens but tolerate self-antigens. T cells recognise foreign antigens through the TCR, a dimeric surface receptor that recognises peptides presented on major histocompatibility complex (MHC) proteins expressed on neighbouring cells. Immature thymocytes express neither the TCR nor its co-receptors CD4 and CD8 and are therefore called double negative (DN) cells (Fig. 10). During so-called β-selection, DN cells undergo rearrangement of the T cell receptor TCRβ gene locus through V(D)J recombination. Only thymocytes that have successfully rearranged their TCR genes and express a functional TCRβ chain that is able to assemble with the TCRα chain and the CD3 co-receptor to form the pre-TCR pass the β-selection checkpoint and undergo pre-TCR-dependent proliferation and differentiation to give rise to double positive (DP) cells that express both CD4 and CD8 receptors as well as the mature TCR. DP cells subsequently undergo so-called positive selection and lineage commitment, whereby the interaction of a DP cell’s TCR with either MHC class I or class II molecules on
the surface of neighbouring stromal cells activates TCR-dependent signalling, which
in turn induces differentiation of the DP cell into a CD8 or CD4 single positive (SP)
cell, depending on the specificity of their TCR (Fig. 10). Nonfunctional DN and DP
cells whose (pre-)TCRs are unable interact with their respective MHC receptor
undergo apoptosis (Ciofani and Zuniga-Pflucker, 2007).

The murine Tox gene was originally identified by Kaye and co-workers in DP
thymocytes activated with phorbol 12-myristate 13-acetate (PMA) and ionomycin,
suggesting a role for TOX in the initial stages of positive selection, consistent with its
high expression in the transitional CD4⁺CD8lo cells (Fig. 10). In contrast to its
expression in activated thymocytes, TOX3 levels in most immature thymocytes were
found to be low. Importantly, upregulation of TOX expression was observed during
positive selection to both CD4 and CD8 single positive cell lineages in wild-type
mice. Surprisingly though, transgenic TOX expression driven by the lymphocyte-
specific lck promoter resulted in an increase of the total number of CD8 SP cells at
the expense of the CD4 SP population (Wilkinson et al., 2002), suggesting different
roles for TOX in the regulation of the CD4 and CD8 thymocyte lineages.

Interestingly, forced TOX expression is able to commit DP cells to differentiate
directly into the CD8 SP phenotype without undergoing intermediate stages observed
in wild-type mice. In a follow-up study, Kaye and colleagues showed that TOX
initiates differentiation of DP thymocytes into CD8 SP, even in the absence of MHC-
dependent positive selection. This upregulation of TOX expression in DP cells is
mediated by the protein phosphatase calcineurin, an important regulator of both CD4⁺
and CD8⁺ lineage commitment. However, these CD8 SP cells are unable to fully
mature and to populate peripheral lymphoid organs. In contrast, the relative number
of CD4 SP cells is decreased during positive selection in TOX transgenic mice
(Aliahmad et al., 2004).

Tox−/− mice show no gross abnormalities outside the immune system although TOX
expression is not restricted to the latter, suggesting that other TOX family members
may be able to compensate for the loss of TOX (Aliahmad and Kaye, 2008).
Furthermore, B lymphocyte development is not affected. In contrast, TOX-deficient
mice suffer from a severe block at the transition from the double dull (DD) to the
CD4⁺8lo stage, a transitional stage at the beginning of positive selection that is
characterised by high TOX expression in wild-type animals. As a consequence, Tox−/−
mice fail to complete the development of all cell types of the CD4⁺ lineage, the most
striking phenotypic effect of TOX deficiency. In contrast, the development of CD8$^+$ cells was not affected by loss of TOX, indicating that the CD4$^{+}$8$^{lo}$ transitional stage is not required for lineage commitment of CD8$^+$ cells. This dependence of CD4 SP differentiation on TOX was likely due to a requirement for MHC class II signalling, which is not necessary for the development of CD8$^+$ cells, and vice versa (Aliahmad and Kaye, 2008).

In addition to its role in the later stages of thymocyte development, TOX plays a role in differentiation of DN to DP cells, as forced TOX expression midway during TCR $\beta$-chain selection is able to induce differentiation, but not proliferation of DP precursor cells (Aliahmad et al., 2004). It is therefore somewhat surprising that $\beta$-selection appears largely uncompromised by loss of TOX (Aliahmad and Kaye, 2008). The findings of Kaye and colleagues do not fully explain the obvious paradox that TOX transgenic mice exhibit an increase of the CD8 SP population at the expense of the number of CD4 SP cells, while knockout of the Tox gene results in a loss of CD4$^+$ cells but has no obvious effect on the development of CD8$^+$ cells. This may be at least partially explained by the interplay between MHC class II-dependent signalling in CD4$^+$ precursor cells, which may be mediated by TOX, while the same may not be the case for MHC class I signalling in immature CD8$^+$ thymocytes. Alternatively, TOX-dependent regulation earlier in the thymocyte lineage, notably during TCR $\beta$-selection may exert its CD4-specific effect not until later during lineage commitment.

In addition to its role in the thymus, two recent studies have demonstrated a function for TOX in the development and differentiation of both mouse and human natural killer cells. In addition, lymphoid tissue organogenesis was inhibited in the absence of TOX (Aliahmad et al., 2010; Yun et al., 2010). In both cases this can be considered analogous to the function of TOX in the T cell lineage: TOX3 influences lineage choice in the innate system, and it does so at times when its expression is highest.
TOX function in thymocyte lineage commitment.

Simplified schematic of thymocyte maturation in the thymus of the mouse. Tox is expressed at distinct stages during β-selection and positive selection, thereby regulating the differentiation of CD4 and CD8 single positive thymocytes. DN, double negative cell; DP, double positive cell; SP, single positive cell; DD, double dull cell; MHC, major histocompatibility complex.

1.7.6 A function for TOX3 in the mammary gland and breast cancer

Considering the role of TOX in thymocyte development, it is intriguing to speculate that TOX3 plays a similar role in the lineage commitment of the mammary gland, in particular in the regulation of luminal cell fate. The complex expression pattern of TOX and the intricate interplay with other factors that regulate cell fate demonstrate that the fine-tuning of TOX expression levels throughout thymocyte development has significant consequences for the development of the individual T cell compartments. By analogy to TOX expression in early CD4⁺CD8lo precursor cells, one could envision that TOX3 expression in bipotent progenitor or luminal progenitor cells induces commitment to the luminal cell lineage. As a consequence, perturbation of TOX3-dependent regulation may result in inappropriate cell fate decisions and ultimately breast tumourigenesis. An example of a transcription factor that plays a role in the differentiation of both the haematopoietic system and the mammary gland is GATA3. Knockout studies have shown that GATA3 is expressed at all stages of thymocyte development, from early progenitors to effector T cells, and like TOX, it is
indispensable for the differentiation of CD4 SP cells (Hosoya et al., 2009; Pai et al., 2003; Ting et al., 1996). Thymocytes from GATA3-deficient mice cannot be redirected to differentiate into the CD8^+ lineage, suggesting that GATA3 expression promotes maintenance or survival of thymocytes that have already chosen the CD4^+ cell fate. Similarly, GATA3 is thought to play a role late in luminal cell differentiation, and in keeping with this notion, re-expression of GATA3 in tumour cells induces their differentiation. Based on the available data it is difficult to predict whether TOX3 acts early or late in luminal cell development and tumourigenesis but one could envision a situation where TOX3 may inhibit activation of a subset of ERα target genes early in the luminal lineage, thus blocking terminal differentiation of ERα-positive cells, while at the same time allowing expansion of progenitors. The resulting cells would be ERα-positive and able to respond to estrogen stimulation but would fail to differentiate, thus giving rise to ERα-positive tumour cells. In this model, selected activation of ERα target genes may be caused by interaction of TOX3 with a distinct subset of ERα co-factors, such as GATA3 or FOXA1, thus committing the epithelial cell to a luminal cell fate. Interestingly, TOX3 has recently been shown cooperate with CITED1 to activate ERE-containing promoters in neurons. However, this transactivation appeared to be independent of E2 stimulation (Dittmer et al., 2010). Despite the lack of a DNA-binding domain CITED1 acts as a transcriptional co-activator in various contexts, notably including ERα-dependent transcription (McBryan et al., 2007; Yahata et al., 2001). CITED1 induces ERα target genes, including amphiregulin, in an E2-dependent manner in the pubertal mouse mammary gland, in particular in luminal cells (McBryan et al., 2007). Furthermore, CITED1 null mice display aberrant mammary gland development during puberty including growth retardation and dilated ducts (Howlin et al., 2006). The expression of both TOX3 and CITED1 is upregulated in the mature luminal mammary epithelial population in human and mice, which in the context of a physical interaction of TOX3 and CITED1 suggests that they may exert their function in concert (Lim et al., 2010). The available data suggest a role for TOX3 in the regulation of luminal mammary epithelial cell fate or differentiation, however the underlying mechanisms are not understood. Moreover, SNPs in the regulatory region of the TOX3 gene are associated with increased breast cancer susceptibility, suggesting that changes in TOX3 expression cause breast cancer by aberrant luminal cell differentiation, but,
once again, the mechanisms are unknown. A recent study showed that the TOX3 gene along with other known developmental regulators selectively lack polycomb repression marks H3K27me3 in pancreatic beta-cells as well as in the cortex and cerebellum or the brain (van Arensbergen et al., 2010). It is tempting to speculate that an inappropriate lack of polycomb repression of the TOX3 gene in cells of the luminal lineage may cause the selective activation of proliferation genes, ultimately leading to transformation.
1.8 Aim of this study and hypothesis

The goal of this PhD project was to characterise new breast cancer genes and ultimately, to create new models for human breast cancer. In this context, I was interested to analyse the function of the putative breast cancer predisposition gene TOX3 in the normal mammary epithelium and breast cancer. Genes that confer susceptibility to breast cancer and/or regulate the mammary epithelial cell are potential targets for therapy and may also serve as predictive markers.

My working hypothesis has been that single nucleotide polymorphisms in the TOX3 regulatory region predispose to luminal breast cancer by increasing the expression of TOX3. This increases the propensity of mammary epithelial cells to differentiate along the luminal lineage rather than become myoepithelial cells but at the same time blocks the final steps of differentiation. At the molecular level TOX3 acts at promoters of estrogen receptor α target genes to recruit factors required for the expression of proliferation genes but inhibits recruitment of factors required for expression of differentiation genes.

Three main objectives evolved for this project. The first was the identification of TOX3 interacting proteins by tandem affinity purification and mass spectrometric analysis. The second was the identification of TOX3 target genes and the signalling pathways that TOX3 is implicated in. The final objective was to create a TOX3 tumour model to test our hypothesis.
2 Materials and Methods

2.1 Cell culture

2.1.1 Preparation of organoids from reduction mammoplasty tissue

The culture of normal breast tissue from reduction mammoplasty samples obtained from Ninewells Hospital, Dundee, UK, was approved by the Tayside Tissue Bank ethics committee, and patients gave informed consent. Patients were healthy, pre-menopausal women between the ages of 21 and 40 years with no previous history of breast cancer. The absence of malignancy was confirmed by histopathological examination of all tissue samples.

In order to obtain organoids from reduction mammoplasty samples, ductal tissue was isolated from mammoplasty samples by a pathologist. The adipose tissue was removed from the fibrous mammary tissue, and the remaining tissue was mechanically dissociated using scalpels. The resulting minced tissue was digested in phenol red-free HEPES-buffered Dulbecco’s Modified Eagle Medium (DMEM)/F12 nutrient mixture (Invitrogen, Cergy Pontoise, France), supplemented with 1 mg/ml collagenase A (Roche Diagnostics, Meylan, France), 100 units/ml penicillin and 100 µg/ml streptomycin (Invitrogen) with agitation at 37°C overnight. The following day, organoids were washed twice in phosphate buffered saline (PBS; Invitrogen) or Hank’s balanced salt solution (HBSS; Invitrogen) and centrifuged to separate fractions containing epithelial cells (500 × g) and fibroblasts (900 × g). Organoids were then either dissociated further to obtain single cells as described below or directly cultured on conventional tissue culture plastic (Thermo Fisher Scientific Nunc, St Herblain, France) or Primaria plastic (BD Biosciences, Le Point de Claix, France). In order to distinguish cell lines derived from different reduction mammoplasty tissue samples, cell lines were designated XS, followed by chronologically increasing numbers. For instance, XS03 denotes reduction mammoplasty number three.
2.1.2 Preparation of single primary mammary epithelial cells

To prepare a single cell suspension, organoids were incubated in 0.25% trypsin (Invitrogen) at 37°C and subsequently mechanically dissociated by pipetting. The enzymatic reaction was stopped with trypsin neutralising solution (Lonza, Cologne, Germany) or 5% heat-inactivated fetal bovine serum (FBS; Pan-Biotech, Aidenbach, Germany) in PBS. Cells were centrifuged at 300 × g and the cell pellet was resuspended in 5 mg/ml warm dispase solution and 1 mg/ml DNase or HyQtase (Thermo Fisher Scientific HyClone) and dissociated further by pipetting. The cell suspension was passed through a 40 µm cell strainer and centrifuged at 300 × g. Single cells were then resuspended in mammosphere medium (MSPM), HMEC medium (HMM+) or WIT medium as described below (Duss et al., 2007; Ince et al., 2007). To distinguish the cells arising from the different culture methods, human mammary epithelial cells grown in HMM+ medium on conventional plastic are in the following referred to as human mammary epithelial cells (HMEC), whereas epithelial cells grown in WIT medium on Primaria dishes were termed breast primary epithelial cells (BPEC) (Ince et al., 2007). All primary cells were cultured at 37°C and 5% CO₂ as well as 5% O₂.

2.1.3 Mammosphere culture of primary mammary epithelial cells

To form mammospheres in suspension, single mammary epithelial cells were seeded at a density of 1x10⁵ cells/ml in suspension on ultra-low attachment dishes (Corning Life Sciences, Lowell, USA) in mammosphere medium (MSPM; phenol red-free HEPES-buffered DMEM/F12 nutrient mixture supplemented with B-27 (Invitrogen) (Brewer et al., 1993), 10 units/ml penicillin and 10 µg/ml streptomycin (Invitrogen), 4 µg/ml heparin (Sigma), 20 ng/ml EGF (Invitrogen), and 20 ng/ml bFGF). HMEC and BPEC maintained as mammospheres were passaged every four to seven days. To this end, mammospheres were centrifuged in a 15-ml conical tube (BD Falcon) at 500 × g, digested enzymatically in 0.05% trypsin-EDTA (Invitrogen) or HyQtase (Thermo Scientific HyClone) and passed through a 40 µm cell strainer (BD Biosciences). Mammospheres were analysed using an Olympus CKX41 microscope and a PixeLINK CMOS camera (PixeLINK, Ottawa, Canada).
2.1.4 Adherent culture of primary mammary epithelial cells

Single mammary epithelial cells were maintained in adherent culture either immediately following dissociation of organoids or following one to several rounds of mammosphere culture. Adherent cells were cultured either on conventional plastic dishes in HMM+ medium (HEPES-buffered DMEM/F12, supplemented with 100 units/ml penicillin and 100 µg/ml streptomycin, 10 mM HEPES (Invitrogen), 5 ng/ml EGF, 10 µg/ml insulin, 20 ng/ml bFGF, 0.5 µg/ml hydrocortisone, 5 µg/ml apo-transferrin, 0.1 µM isoproterenol, 1 µM ethanolamine, 1 µM O-phosphoethanolamine and, in the case of ESR1-expressing cell lines, 1 nM β-estradiol) or in WIT medium for primary cells as described by Ince and colleagues (Ince et al., 2007) on Primaria dishes (BD Biosciences). All chemicals were purchased from Sigma (Sigma-Aldrich, Lyon, France) unless otherwise noted. Cells were passaged every three to four days and maintained in filtered conditioned medium. Dissociated cells were seeded at a density of $10^5$ cells/ml.

2.1.5 Isolation and culture of primary mammary fibroblasts

Single primary human mammary fibroblasts (HMF) were isolated from reduction mammoplasty tissue samples as described above for epithelial cells. HMF were separated from epithelial cells by centrifugation at the organoid stage and then processed separately. HMF were cultured on conventional plastic dishes in RPMI-1640 medium (Invitrogen), supplemented with 100 units/ml penicillin and 100 µg/ml streptomycin and 10-20% FBS at high density but without being permitted to reach confluency.

2.1.6 Culture of established tumour cell lines

The cell lines ZR-75-1 and CAMA-1 were a kind gift from Françoise Bonnet, Institut Bergonié, Bordeaux, France, and Paul Reynolds, University of St Andrews, UK, respectively. All other established cell lines were obtained from the American Type Culture Collection (ATCC). Human breast cancer cell lines MCF-7, T-47D, BT-474, and ZR-75-1 were maintained in RMPI-1640 medium supplemented with 100
units/ml penicillin and 100 µg/ml streptomycin and 10-20% FBS. Human breast cancer cell lines CAMA-1, MDA-MB-231, MDA-MB-361, MDA-MB-415, and UACC-812, and human embryonic kidney cells transformed with SV40 large-T antigen (HEK 293T) were maintained in DMEM supplemented with 100 units/ml penicillin and 100 µg/ml streptomycin and 10-20% FBS. MCF-10A cells were maintained in HEPES-buffered DMEM/F12, supplemented with 100 units/ml penicillin and 100 µg/ml streptomycin, 10% FBS, 20 ng/ml EGF, 10 µg/ml insulin, 0.5 µg/ml hydrocortisone, and 100 ng/ml cholera toxin (Sigma). The transformed HMEC line AJ4 expressing ESR1, BMI1, TERT, and MYC and enhanced cyan fluorescent protein (ECFP) has been described previously and was maintained in HMM+ medium (Duss et al., 2007).

2.1.7 Colony forming cell (CFC) assay

CFC assays to assess progenitor properties of primary mammary epithelial cells and MCF-10A cells were performed essentially as described (Bachelard-Cascales et al., 2010). Briefly, 10⁵ cells/ml unsorted single primary mammary cells from freshly dissociated organoids or MCF-10A cells were seeded onto adherent irradiated mouse embryonic fibroblasts (MEF) in the respective culture medium. Colony formation, phenotype and size were assessed at different time points by fixation of the cells in ice-cold methanol and subsequent staining with May-Grünwald colorant.

2.1.8 Terminal ductal lobular unit (TDLU) assay

The TDLU assay was used as a 3D in vitro stem cell assay and was performed as described (Bachelard-Cascales et al., 2010). In brief, Lab-Tek chamber slides (Thermo Scientific Nunc) were filled growth factor-reduced matrigel (BD Biosciences). For primary cells, matrigel was mixed with irradiated MEF feeder cells. For established tumour cell lines, matrigel alone was used. Matrigel was allowed to set briefly and 5x10³ cells resuspended in 2 µl PBS were injected into the middle of the semi-hardened matrigel. Lab-Tek chambers were immediately placed at 37°C to cause matrigel to harden and in order to prevent diffusion of the injected cells. Matrigel was subsequently covered with the respective medium. After 15-20 days,
Materials and Methods

epithelial outgrowths were paraffin-embedded, sectioned, and haematoxylin and eosin (H&E)-stained for histological analysis.

2.1.9 Flow cytometry

For phenotypic analysis of epithelial subpopulations, cells were resuspended in HBSS, supplemented with 2% FBS and incubated for 30 minutes to 1 hour with 0.1 μg of the following antibodies per 10^6 cells: IgG1-PE-Cy5, CD10 PE-Cy7, EpCAM/CD326-APC, CD29-PeCy5, CD24-PE, CD49f-PE, CD49d-PE (all BD Biosciences). Cells were centrifuged, washed and fixed in 2% paraformaldehyde (Sigma) in PBS. Fluorescence-activated cell sorting (FACS) analysis was performed using a FACSCalibur flow cytometer (BD Biosciences).

2.1.10 Production of lentiviral particles

Lentiviral particles were produced by calcium phosphate transfection of HEK 293T cells using a second-generation lentiviral packaging system according to standard techniques (Dull et al., 1998). Briefly, a total of 4.5 x 10^6 cells were seeded in standard medium on a 10 cm dish 24 hours prior to transfection. One hour prior to transfection, the medium was replaced with 4.5 ml standard medium, supplemented with 25 μM chloroquine. 15-20 μg of lentiviral expression plasmid DNA, 4 μg of vesicular stomatitis virus glycoprotein (VSV-G) expression plasmid (pMD2.G) and 10 μg of gag-pol packaging plasmid (PCMVdR8.74) were diluted with 25 mM HEPES to a final volume of 250 μl (Dull et al., 1998; Follenzi and Naldini, 2002). After the addition of 250 μl of 0.5 M CaCl_2, the DNA mix was added drop-wise to 500 μl 2x HEPES-buffered saline (Sigma) while vortexting. The transfection mix was incubated for 20 minutes and added drop-wise to the cells. 6 hours post transfection, the medium was replaced with 5 ml of fresh medium. The supernatant containing the lentiviral particles was collected 48 hours post transfection and filtered using a 22-μm mixed cellulose ester filter (Millipore) and stored at -80°C. Lentiviral titres were determined for each viral batch by serial dilution infections of MCF-7 cells or primary mammary cells and subsequent puromycin treatment or detection of marker gene
fluorescence, according to the viral construct. Typical viral titres obtained were $10^7$ infectious units/ml.

2.1.11 Lentiviral infection

For lentiviral infection of mammosphere-forming HMEC and BPEC, cells isolated from freshly dissociated organoids were allowed to recover in suspension overnight. Primary cells were infected in suspension for 6 hours at a multiplicity of infection (MOI) of 5-50, depending on the lentiviral construct and experiment, and seeded on ultra low attachment dishes at a density of $10^5$ cells/ml. Infections were carried out in the respective culture medium supplemented with 8 $\mu$g/ml polybrene (Sigma). Alternatively, single HMEC and BPEC were seeded in adherent culture and infected 24 hours later as described for cells in suspension. Established breast cancer cell lines were infected essentially as described for HMEC and BPEC in adherent culture with the modification that the infection was allowed to take place overnight at a MOI of 5-20. 48 hours post infection, transduced cells were selected by puromycin treatment for 72 hours. Alternatively, successful transduction was confirmed by the presence of a fluorescent marker gene. For infection with multiple viruses, cells were subjected to several rounds of infection, and successful transduction was confirmed after each round.

2.1.12 Dual-luciferase assay

To analyse the effect of TOX3 expression on the transcriptional activity of minimal promoters containing estrogen receptor response elements (ERE), MCF-7 and T-47D cells were transiently transfected in triplicate using ExGen 500 polyethylenimine transfection reagent (Euromedex) according to the manufacturer’s recommendations. Briefly, $3\times10^5$ cells per well were seeded on a 12-well plate 24 hours prior to transfection. For dual-luciferase reporter assays, 225-400 ng of firefly luciferase control vector pGL3-basic (Clontech, Saint-Germain-en-Laye, France) or the reporter plasmids pGL3-pS2-luc or pGL3-ERE-TK-luc were co-transfected with 50-225 ng lentiviral TOX3 or DsRed2 control expression plasmid. The firefly luciferase reporter plasmids pGL3-pS2-luc containing the proximal pS2 (trefoil factor 1, TFF1) promoter
and pGL3-ERE-TK-luc containing the Vitellogenin ERE along with a minimal thymidine kinase promoter were a kind gift from George Reid, EMBL, Heidelberg, Germany. 50 ng of the pRL-TK or pRL-SV40 Renilla luciferase reporter plasmids (Clontech) were co-transfected as an internal control. Lentiviral expression plasmids are described below. 24 hours post transfection, cells were treated with 100 nM 17-beta-estradiol (Sigma) or 1 µM fulvestrant/ICI 182,780 (Tocris Bioscience, Bristol, UK) and subsequently incubated for an additional 24 hours. Preliminary experiments had shown that cells maintained in phenol red-free RPMI-1640 medium (Invitrogen) supplemented with 5% charcoal dextran-treated FBS (Invitrogen) to exclude potential estrogenic effects of phenol-red displayed an abnormal morphology. I therefore chose to use the estrogen receptor antagonist fulvestrant in complete medium instead of an untreated control. Cell lysates were prepared and the luciferase assay was carried out using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer’s recommendations on a LUMIstar Galaxy luminometer (BMG Labtechnologies, Offenburg, Germany). Data were analysed using Microsoft Excel software.

2.2 Xenografts

Animal experiments were authorised by the United Kingdom Home Office and the French Ministère de l’alimentation, de l’agriculture et de la pêche. For xenograft experiments using transformed primary cells, 75,000 HMEC AJ4, BPEC XS08.4 or BPEC XS08.ctrl and 25,000 HMF were mixed with 12.5 vol% Matrigel (BD Biosciences) and injected into the fourth mammary gland of 30 week-old female nonobese diabetic/Cg-Prkdc<sup>scid</sup> Il2rg<sup>tm1Wjl</sup>/SzJ mice (commonly know as NOD scid gamma or NSG) mice (Jackson Laboratory, Bar Harbor, USA). NSG mice are severely immunocompromised due to the absence of mature T or B cells, as well as lack of functional NK cells and deficiency in cytokine signalling. One week prior to the injection of human mammary cells, 0.5 mg beta-estradiol slow-release silicon pellets were inserted subcutaneously into the neck region of animals as described (Duss et al., 2007). Mice were sacrificed and examined for tumour and metastasis incidence at different time points using a Leica MZ FLIII fluorescence stereomicroscope (Leica, Wetzlar, Germany) and a PixeLINK CMOS camera.
Tumours were paraffin-embedded and sectioned. H&E staining was carried out using standard methods.

For MCF-7 xenografts, 100,000 cells transduced with hPGK-TOX3 (pXS-36) lentiviral vector or a luciferase control vector (pER-6) at a MOI of 20, were injected. Cells had additionally been transduced with a hPGK-GFP vector (pXS-17) for tracking of cells in the mouse. Cells were grafted by intraductual injection without supplementation of matrigel or HMF in 6-8-week old female NSG mice as previously described (Behbod et al., 2009). Beta-estradiol pellets were inserted on the day of xenografts. Tumour formation was monitored by in vivo fluorescence imaging using a Photon Imager with M3 Vision software (Biospace Labs, Paris, France) for 21 days, and mice were sacrificed and examined for tumour incidence at day 7 and 21 post injection.

2.3 Lentiviral vectors

Lentiviral expression vectors were created by Gateway cloning (Invitrogen) of the open reading frame (ORF) of interest into one of three different lentiviral destination vectors, which provided the backbone for the viral expression plasmid. The lentiviral destination vector pSD-69 has been described previously (Duss et al., 2007). It contains the human phosphoglycerate kinase (PGK) promoter to drive the expression of the gene of interest adjacent to a Gateway attR cassette as well as the puromycin acetyltransferase gene under the control of the murine PGK promoter, cloned into the pRRLhPGK.GFP.SIN18 backbone (Dull et al., 1998). The lentiviral destination vectors pJH-3982 and pJH-4282 were a kind gift from J. Huelsken, ISREC, Switzerland. pJH-3982 contains the human ubiquitin C promoter (UbiC) to drive the expression of the gene of interest, an attR cassette, as well as the enhanced GFP (EGFP) gene under the control of the human hPGK promoter. The lentiviral destination vector pJH-4282 contains the CMV promoter upstream of an attR cassette as well as the marker gene EGFP under the control of the human PGK promoter.
2.3.1 TOX3 lentiviral expression vectors

The TOX3 ORF was transferred from the TOX3 UltimateORF clone IOH61112 (corresponding to transcript variant 1, accession NM_001080430) pDONR entry plasmid (Invitrogen) into pSD-69 and pJH-3982 by Gateway LR recombination (Invitrogen) to create the lentiviral expression vectors pXS-36 and pXS-39, respectively. Lentiviral vectors derived from pSD-69 used as a control encoded β-glucuronidase (gusA, GUS; pSD-86) or the fluorescent marker genes green fluorescent protein (GFP; pER-15) or DsRed2 (pSD-136) (Duss et al., 2007). The control vector derived from pJH-3982 contained the DsRed2 ORF (pXS-41). Lentiviral vectors were amplified using electrocompetent *E. coli* strains Stbl2 or Stbl3, and plasmid DNA was purified using Qiagen Plasmid Kits (Qiagen, Courtaboeuf, France) according to the manufacturer’s instructions.

2.3.2 NTAP-TOX3 lentiviral expression vectors

Lentiviral expression constructs for tandem affinity purification (TAP) of TOX3 and its interacting proteins were created by a splice PCR-based strategy combined with Gateway cloning (Invitrogen). In the first two PCR steps, the TOX3 open reading frame (transcript variant 1, accession NM_001080430) was amplified from the Ultimate ORF clone IOH61112 (Invitrogen), while the TAP-tag was amplified from pNTAP (Agilent Technologies, Massy, France). At the same time, a linker region was introduced to allow the fusion of the sequences in a third overlap PCR. The primers (Eurogentec, Angers, France) are listed in Table 3. The final PCR product was flanked by *attB* sites and was cloned into the pDONR201 entry vector (Invitrogen) by Gateway BP recombination. The N-terminally TAP-tagged TOX3 ORF was subsequently cloned into the pSD-69, pJH-3982 and pJH-4282 lentiviral destination vectors by Gateway LR recombination to give pXS-95, pXS-96, and pXS-85, respectively. The corresponding control lentiviral vectors pXS-98, pXS-99, and pXS-72 containing the TAP-tag and a linker region only were cloned accordingly. Entry and expression vectors were confirmed by sequencing.
Table 3. PCR primers used for TAP-TOX3 and TAP control cloning

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>oXS6 (attB1_TAP_fwd)</td>
<td>GGGGACAAGTTTGTACAAGAACAGCGCTGCCACCATGAAGCGACG</td>
</tr>
<tr>
<td></td>
<td>ATGGAAAAAGAATTTC</td>
</tr>
<tr>
<td>oXS7 (TAP_linker_rev)</td>
<td>TCCACCGGTACCTCTGAAACGCTGGTACCTTTGAGGTACCGGGAGAGTG</td>
</tr>
<tr>
<td>oXS8 (NTAP-TOX3_fwd)</td>
<td>GCTGCAAGCTGGCCGCTGTACCGGTACCGGGAGATGTTGAGGTTTCTACC</td>
</tr>
<tr>
<td>oXS9 (TOX3_attB2_rev)</td>
<td>GGGGACCACTTTGTGACAAAGAGCTGGTGCTGCTAGAAATACTGACCTCGGAT</td>
</tr>
<tr>
<td>oXS10 (attB1)</td>
<td>GGGGACAAATTTGTAACAAAAAGCGACGC</td>
</tr>
<tr>
<td>oXS11 (attB2)</td>
<td>GGGGACCACTTTGTGACAAAGAGCTGGT</td>
</tr>
<tr>
<td>oXS13 (TAPtag_control_rev)</td>
<td>GGGGACCACTTTGTGACAAAGAGCTGGTCTAGCTGTCGAGAAAGCTTG</td>
</tr>
<tr>
<td>(TAPtag_control_rev)</td>
<td>GATCCTCCACCGGTACCCTCTGAC</td>
</tr>
</tbody>
</table>

2.3.3 TOX3-GFP lentiviral expression vector

A C-terminal TOX3-GFP lentiviral fusion construct was created by a splice PCR-based strategy combined with Gateway cloning (Invitrogen). In the first two PCR steps, the TOX3 open reading frame was amplified from the Ultimate ORF clone IOH61112 (Invitrogen), while EGFP was amplified from pEGFP-N1 (Clontech). The complete fusion TOX3-GFP ORF was created by a third overlap PCR. The PCR primers (Eurogentec) are listed in Table 4. The final PCR product contained flanking attB sites and was cloned into the pDONR201 entry vector (Invitrogen) by Gateway BP recombination. The TOX3-GFP fusion ORF was subsequently cloned into the pSD-69 lentiviral destination vector by Gateway LR recombination to give pXS-71. Entry and expression vectors were confirmed by sequencing.

Table 4. PCR primers used for TOX3-GFP cloning

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>oXS14 (attB1_TOX3_fwd)</td>
<td>GGGGACAAGTTTGTACAAGAACAGCGCTGCCACCATGAAGCGACG</td>
</tr>
<tr>
<td></td>
<td>AGGTTTCTACCCGCGG</td>
</tr>
<tr>
<td>oXS15 (TOX3_MCS_rev)</td>
<td>GGGGACCACTTTGTGACAAAGAGCTGGTCTAGCTGTCGAGAATACCTT</td>
</tr>
<tr>
<td>oXS16 (MCS_EGFP_fwd)</td>
<td>GCCCGGGATACCCGCGTGATCCGCCACCATGAGCGAGGGCGGAGAGCTG</td>
</tr>
<tr>
<td>oXS17 (EGFP_attB2_rev)</td>
<td>GGGGACCACTTTGTACAGAAAGCGCTGGTGACCTCGCCGCGGCTTACT</td>
</tr>
<tr>
<td></td>
<td>TGTGACGCTGCGGATG</td>
</tr>
</tbody>
</table>
2.3.4 HA-TOX3 lentiviral expression vector

To create an N-terminal haemagglutinin (HA)-tagged TOX3 lentiviral expression vector, the TOX3 ORF was PCR-amplified from the pXS-36 (hPGK-TOX3) plasmid using the primers given in Table 5. The forward primer comprised the HA sequence. The PCR product was cloned into the pDONR201 entry vector (Invitrogen) by Gateway BP recombination via flanking attB sites, and the insert was subsequently transferred into the pSD-69 lentiviral hPGK destination vector by LR recombination to give the HA-TOX3 lentiviral expression vector pER-47.

Table 5. PCR primers used for HA-TOX3 cloning

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>oER20 (attB1_HA-TOX3 fwd)</td>
<td>GGCAAGTTTTGTACAAAAAAGCAGGCCACCATGTACCCATACTGAGATGTCCA</td>
</tr>
<tr>
<td></td>
<td>GATTACGCTGCGGATCCATGGATGTGAGTTCTACCCGCGGCGGCG</td>
</tr>
<tr>
<td>oER21 (TOX3_attB2_rev)</td>
<td>CCACCTTTGTACAAGAAAGCTGGGTCTAGAGTTACTGAGAACATGACCTGC</td>
</tr>
<tr>
<td></td>
<td>GATAACTTTGAGTCTGTGCTGAG</td>
</tr>
</tbody>
</table>

2.3.5 TOX3 lentiviral small hairpin microRNA vectors

A set of two pGIPZ lentiviral small hairpin miRNA constructs directed against sequences in the 3' untranslated region of TOX3 (pXS-88 and pXS-89) and a scrambled pGIPZ control vector (pXS-62) were purchased from Open Biosystems (Huntsville, USA).

2.3.6 Other lentiviral vectors

The GFP-expressing lentiviral vectors pRRLsin.PPTs.hCMV.GFPpre (pXS-3) and pRRLsin.PPTs.hPGK.GFPpre (pXS-17) used to fluorescently label cells for xenografts without conferring antibiotic resistance were a kind gift from L. Naldini, San Raffaele Scientific Institute, Italy. pRLLsin.cPPT.MND.LUC.wpre (pER-6) pRLLsin.cPPT.MND.Tomato.wpre (pER-8) expressing luciferase and Tomato, respectively, under the control of the synthetic MND promoter were generously provided by H. de Verneuil, U876 INSERM, Bordeaux, France. The MND promoter comprises the U3 region of a modified Moloney murine leukemia virus (MoMuLV) LTR with myeloproliferative sarcoma virus enhancer (Li et al., 2010). The pSD-69-
derived lentiviral expression vectors coding for ESR1 (pSD-82), TERT (pSD-83), BMI1 (pSD-84), and MYC (pSD-94) have been described previously (Duss et al., 2007). A complete list of plasmids used in this study is included in the Appendix.

2.4 RNA expression analysis

2.4.1 Isolation and reverse transcription of RNA

Total RNA was isolated using the Qiagen RNeasy Kit (Qiagen) according to the manufacturer’s recommendations. The integrity of RNA for subsequent microarray analysis was confirmed using the Agilent Bioanalyzer system (Agilent Technologies, Massy, France). RNA was quantified using a NanoDrop spectrophotometer (Thermo Scientific NanoDrop, Wilmington, USA). 100 ng-1 µg of total RNA were reverse transcribed using random hexamers and SuperScript II reverse transcriptase (Invitrogen).

2.4.2 Quantitative PCR

Following reverse transcription, cDNA was amplified by quantitative PCR (qPCR) using the Power SYBR Green PCR Master Mix and the primers detailed in Table 6 on the GeneAmp PCR System 9700 (both Applied Biosystems, Courtaboeuf, France). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was selected as reference gene by geNorm analysis (Vandesompele et al., 2002).

Table 6. Primers used for quantitative PCR

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>oXS20</td>
<td>ACATCGCTCAGACACCATGGGGA</td>
</tr>
<tr>
<td>GAPDH_qPCR_fwd</td>
<td>CCAGGGCGCCAAATACGACCCAAA</td>
</tr>
<tr>
<td>oXS23</td>
<td>CATACAGGGCCAGCTCGTCTTAAG</td>
</tr>
<tr>
<td>GAPDH_qPCR_rev</td>
<td>TGAACAGAAACGGATGTTCCTGGGCTT</td>
</tr>
<tr>
<td>oXS24</td>
<td>ACATCGCTCAGACACCATGGGGA</td>
</tr>
<tr>
<td>TOX3_qPCR_fwd</td>
<td>CCAGGGCGCCAAATACGACCCAAA</td>
</tr>
<tr>
<td>oXS25</td>
<td>CATACAGGGCCAGCTCGTCTTAAG</td>
</tr>
<tr>
<td>TOX3_qPCR_rev</td>
<td>TGAACAGAAACGGATGTTCCTGGGCTT</td>
</tr>
</tbody>
</table>
2.4.3 Microarray analysis

Gene expression analysis of 162 primary pre-treatment tumour biopsies and 3 samples of normal tissue from the neoadjuvant clinical trial EORTC 10994/BIG 00-01 using Affymetrix GeneChip Human X3P Arrays has been described previously (Farmer et al., 2009; Farmer et al., 2005). Microarray analysis of TOX3 target genes was performed using Illumina HumanHT-12 Expression BeadChips (IntegraGen, Evry, France). Data were processed as described below.

2.5 Protein expression analysis

2.5.1 TOX3 antibodies

A rabbit polyclonal peptide antibody raised against amino acids NEEDADEANR (aa 217-226) contained in both isoforms of human TOX3 was custom-made (Abgent, Milton Park, UK). Two rabbits were immunised with the peptide conjugated with KLH (Keyhole limpet hemocyanin) in six boosts with one week between individual boosts. The corresponding protein A-purified sera were termed TOX3 Rb57 and TOX3 Rb58, respectively. TOX3 Rb57 was further affinity-purified as described below. A goat polyclonal peptide TOX3 antibody (ab77432) raised against amino acids AGDPASLDFAQC was purchased from Abcam (Abcam, Cambridge, UK). The latter peptide sequence corresponds to N-terminal amino acids 10-21 of human TOX3 isoform 1 and is not contained in isoform 2.

2.5.2 Preparation of whole cell extracts and protein quantitation

Cells were washed twice in cold PBS and whole cell extracts (whole cell lysates) for Western blot analysis were prepared by direct lysis of cells in 2x protein loading buffer (4% SDS, 20% glycerol, 0.004% bromophenol blue, 160 mM Tris-HCl, pH 6.8, 200 mM DTT). Equal protein loading was confirmed by Coomassie blue staining of a separate SDS-PAGE gel prior to Western blot analysis and by Ponceau S (Sigma) staining of nitrocellulose membranes following transfer. For TAP purification and co-immunoprecipitation, cells were lysed in TAP lysis buffer, RIPA buffer, modified RIPA buffer or NET buffer, depending on the experiment, as described below. The
protein concentration was then determined using the Bio-Rad DC Protein Assay (Bio-Rad, Hercules, USA) according to the manufacturer’s recommendations.

2.5.3 Western blot analysis

Western blotting was carried out according to standard techniques. Antibodies were diluted in 5% milk in PBS-Tween 20 (Sigma). The following other antibodies were used: TOX3 antibodies as described above, AGR2 (Abnova, Taipei City, Taiwan), BMI1 (Millipore), CBP epitope tag (Millipore), CITED1 (Abcam), ESR1 (Labvision Thermo Fisher Scientific, Fremont, USA), GFP (Santa Cruz Biotechnologies, Santa Cruz, USA), HA probe (Santa Cruz Biotechnologies, Santa Cruz, USA), HSP70 (Stressgen, Enzo Life Sciences, Villeurbanne, France), IRS1 (Millipore), p16 (Novocastra, Newcastle, UK), p21 (Santa Cruz Biotechnologies), p53 DO1 (kind gift from David Lane, University of Dundee, UK), PGR isoforms A and B (Dako, Trappes, France), α-tubulin (Sigma), YBX1 (Abcam), donkey secondary antibodies against rabbit and mouse IgG (GE Healthcare, Orsay, France). A comprehensive list of antibody clones and working concentrations is given in the Appendix.

For peptide blocking experiments, TOX3 immunising peptide (Abgent) was added to the antibody solution at a concentration that was two-fold higher by mass (~150-fold molar excess) than the concentration of the respective TOX3 antibody.

2.5.4 Co-immunoprecipitation

For immunoprecipitation of TOX3 and interacting proteins, HEK 293T were transfected using calcium phosphate as described above, MCF-7 and BT-474 cells were transduced with lentiviruses. $10^7$ cells were lysed in 500 µl of RIPA buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 1% SDS, 1 mM EDTA) or modified RIPA buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% NP-40, 0.25% sodium deoxycholate, 1 mM EDTA), supplemented with cOmplete protease inhibitor cocktail (Roche Diagnostics, Meylan, France) and 1 mM PMSF. Lysates were diluted two- to ten-fold in NET buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 1% Nonidet-P40, 5 mM EDTA), supplemented with protease inhibitors and incubated with 1-2 µg of antibody. Antibody-antigen complexes were
subsequently incubated with protein A- or protein G-coupled Dynabeads (Invitrogen). Protein complexes were eluted in 2x protein loading buffer and analysed by Western blotting.

2.5.5 Immunofluorescent staining

For indirect immunofluorescent staining, established cell lines were seeded on glass coverslips. HEK 293T cells were calcium phosphate transfected as described above. MCF-7 cells were either transfected using FuGENE HD transfection reagent (Roche Diagnostics) according to the manufacturer’s recommendations or transduced with lentiviruses. Cells were fixed in 3.7% formaldehyde in PBS, incubated in 50 mM NH₄Cl, permeabilised in 0.5% Triton X-100 and subsequently blocked in 0.2% gelatine. Cells were incubated with primary antibody overnight at 4ºC. The following day, cells were washed and incubated with fluorophore-conjugated secondary antibody for 1 hour at room temperature in the dark, followed by staining with 100 ng/ml DAPI (Sigma) in PBS. Primary HMEC and BPEC were grown on conventional cell culture plastic or Primaria plastic, respectively, and fixed using a 1:1 mix of ice-cold methanol and acetone. Subsequent immunofluorescent staining of primary cells was carried out directly on cell culture dishes as described (Duss et al., 2007). The following primary antibodies were used for indirect immunofluorescence: cytokeratin 14, cytokeratin 18 (both kind gifts from Birgit Lane, University of Dundee, UK), and γ-H2AX (Millipore). Secondary antibodies used were: Alexa 488- or 568-conjugated anti-rabbit or anti-mouse (Invitrogen Molecular Probes).

For immunofluorescent analysis of TOX3-GFP fusion protein, HEK 293T cells were seeded on polylysine-covered glass coverslips (Sigma) and calcium phosphate-transfected as described above. MCF-7 cells were transduced with the TOX3-GFP lentivirus and seeded on coverslips. Confluent cells were washed twice with cold PBS and fixed using a 1:1 mix of methanol and acetone. Fixed cells were allowed to dry before staining with 100 ng/ml DAPI in PBS.

For confocal microscopy, coverslips were mounted onto microscope slides using 10 µl of Vectashield mounting medium (Vector Laboratories, Burlingame, USA) or 5 µl of Fluoromount-G mounting medium (SouthernBiotech, Birmingham, USA) and
Materials and Methods

sealed with nail polish. Slides were allowed to dry at 4°C overnight. Analysis of stained cells was performed using a Zeiss LSM 510 confocal microscope.

2.5.6 Immunohistochemistry

Tissue microarrays were prepared as previously described (Banneau et al., 2010). Haematoxylin and eosin staining was on paraffin-embedded cell lines and tissue microarrays was carried out according to standard protocols. Immunohistochemical staining of paraffin-embedded cell lines and tissue microarrays was carried out by the pathology service at Institut Bergonié, Bordeaux using the Ventana ultraView Universal DAB Detection Kit and the automated slide stainer Ventana BenchMark XT (Ventana Medical Systems, Tuscon, USA) according to the manufacturer’s instructions. Staining conditions were optimised for each antibody: TOX3 (Abcam), TOX3 (Abgent), ERα, cytokeratin 14 (both Dako, Trappes, France). Stained sections and tissue microarrays were analysed using a Leica Leitz DMRB microscope with a Nikon DXM1200 camera.

2.6 Tandem affinity purification (TAP) of TOX3

2.6.1 TAP

Tandem affinity purification (TAP) of TOX3 and interacting proteins was performed using the InterPlay Mammalian TAP Kit (Agilent Technologies) according to the manufacturer’s recommendations with the following changes made after optimisation of the protocol as described in the Results chapter. Briefly, HEK 293T cells were calcium phosphate-transfected with the NTAP-TOX3 lentiviral construct or the corresponding control construct as described above. 48 hours post transfection, 10^8 transfected cells per condition were washed twice in cold PBS and scraped in PBS. Alternatively, 10^8 lentiviral-transduced HEK 293T cells were used per condition. All following steps were carried out at 4°C, and all washes and incubation steps were performed with agitation of the sample. Cells were pelleted, and cell pellets were resuspended in 10 ml TAP lysis buffer (10 mM Tris-HCl pH 8.0, 400 mM NaCl, 0.1% Nonidet-P40, 2 mM EDTA, 10 mM 2-mercaptoethanol), supplemented with
Materials and Methods

cOmplete protease and PhosSTOP phosphatase inhibitor cocktails (Roche Diagnostics) and 1 mM PMSF. Cells were lysed by three freeze-thaw cycles at -80°C and centrifuged at 13,000 rpm in a microcentrifuge. The supernatant containing soluble protein was incubated with the streptavidin resin for 2 hours. Following three washes in 1 ml of streptavidin binding buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Nonidet-P40, 2 mM EDTA, 10 mM 2-mercaptoethanol) supplemented with all inhibitors, the TAP-tagged protein was eluted from the streptavidin resin by incubation in streptavidin elution buffer (10 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.1% Nonidet-P40, 2 mM biotin, 10 mM 2-mercaptoethanol) supplemented with all inhibitors protected from light for 1 hour. 1 ml of streptavidin eluate was supplemented with 20 µl streptavidin supernatant supplement at a final concentration of 50 mM magnesium acetate, 50 mM imidazole, 100 mM CaCl₂. 1 ml of eluate was incubated with the calmodulin resin in 4 ml of calmodulin binding buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Nonidet-P40, 1 mM magnesium acetate, 1 mM imidazole, 2 mM CaCl₂, 10 mM 2-mercaptoethanol) supplemented with all inhibitors for 2 hours to overnight. Following three washes in calmodulin binding buffer, the TAP-tagged protein was eluted from the calmodulin resin by incubation in 400-500 µl calmodulin elution buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Nonidet-P40, 1 mM magnesium acetate, 1 mM imidazole, 10 mM EGTA, 10 mM 2-mercaptoethanol) for 1 hour. For silver stain and Western blot analysis, eluted protein was concentrated with trichloroacetic acid (TCA). To this end, TCA was added to the eluate to a final concentration of 15% (v/v). The solution was vortexed, incubated at 4°C for 30 minutes and centrifuged at 13,000 × g at 4°C for 15 minutes in a microcentrifuge. The supernatant was carefully aspirated, and the precipitate was washed two times with 1 ml of ice-cold acetone and centrifuged at 13,000 g at 4°C for 10 minutes. The pellet was allowed to air-dry briefly and then resuspended in an appropriate volume of calmodulin elution buffer. For mass spectrometric analysis, eluted protein was concentrated using an Amicon Ultra-0.5 centrifugal filter device with a 10,000 nominal molecular weight limit (Millipore).
2.6.2 Optimisation of TAP conditions

For optimisation of cell lysis, the following alternative lysis buffers were tested: KCl lysis buffer (20 mM Tris-HCl, pH 8.0, 400 mM KCl, 0.1% NP-40), which was modified from Dignam and Roeder buffers B and C for the preparation of nuclear extracts (Dignam et al., 1983), or NET buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% NP-40, 5 mM EDTA). To optimise the streptavidin elution efficiency, SEB was supplemented with 400 mM NaCl or 4 mM biotin, alternatively. To block non-specific binding of chromatin to the streptavidin beads, which could reduce the binding capacity of the resin for NTAP-TOX3, cell lysates were supplemented with 100 µg/ml salmon sperm DNA.

2.6.3 Silver stain analysis

Proteins present in the individual fractions after each purification step were analysed by silver stain analysis using the Silver Stain Plus kit (Bio-Rad Laboratories) according to the manufacturer’s recommendations.

2.6.4 Mass spectrometry

For mass spectrometric analysis of purified NTAP-TOX3 and interacting proteins, concentrated protein was separated on a NuPAGE Bis-Tris 4-12% gradient gel (Invitrogen) and stained with Coomassie Brilliant Blue G-250 (colloidal blue; 30 g ammonium sulfate, 0.2 g Coomassie Brilliant Blue G-250, 6 ml orthophosphoric acid, 40 ml ethanol, H₂O ad 200 ml). The gel lane was separated into 16 sections. The protein was trypsin-digested, extracted from the gel and subsequently subjected to nano-scale liquid chromatographic tandem mass spectrometry (nLC-MS/MS) using a LTQ OrbiTrap spectrometer (Thermo Scientific). Identified peptide sequences were compared to the Homo sapiens UniProtKB/Swiss-Prot database http://www.expasy.org/sprot/. A detailed description of the mass spectrometric analysis is included in the Appendix.
2.6.5  Affinity purification of TOX3 antibody

Rabbit polyclonal TOX3 Rb57 antibody was affinity-purified using NTAP-TOX3 protein immobilised on streptavidin-coated magnetic beads. To this end, HEK 293T cells were transfected with the hPGK-NTAP-TOX3 lentiviral construct or the control construct, respectively, and cell lysates were prepared in TAP lysis buffer containing 400 mM NaCl and protease inhibitors as described above. Following centrifugation to remove the cell debris, 5mM EDTA was added to the supernatant. Dynabeads MyOne Streptavidin C1 (Invitrogen) were washed three times in TAP lysis buffer supplemented with 5 mM EDTA and subsequently incubated with NTAP-TOX3 lysate or NTAP control supernatant, respectively, for 1 hour at 4°C. Beads were pulled down with the magnet and washed once with TAP lysis buffer containing 400 mM NaCl, 5 mM EDTA, followed by two more washes in TAP lysis buffer containing 150 mM NaCl, 5 mM EDTA. TAP-tagged TOX3 was then cross-linked to the streptavidin-coated Dynabeads using 20 mM dimethyl pimelimidate (DMP) in 0.2 M sodium borate, pH 9.0 for 30 minutes at room temperature. The reaction was stopped by incubation of the beads in 0.2 M ethanolamine, pH 8.0. Unbound NTAP-TOX3 was removed by washing in 0.1 M glycine-HCl, pH 2.5. NTAP-TOX3-coated Dynabeads were subsequently incubated with TOX3 Rb57 antibody overnight at 4°C. The following day, Dynabeads-NTAP-TOX3-antibody complexes were washed three times with 10 mM Tris-HCl pH 7.4, 150 mM NaCl. Purified TOX3 antibody (TOX3 Rb57P) was eluted in five fractions using 200 µl of 0.1 M glycine-HCl, pH 2.5. To prevent denaturation of the antibody, elution fractions were neutralised immediately by addition of 50 µl of 1 M Tris-HCl pH 8.0. To remove antibody that bound non-specifically to the TAP tag, the purified antibody fractions were incubated with TAP control-coated Dynabeads. Aliquots of each purification step were analysed by SDS-PAGE and Coomassie Brilliant Blue staining, and the antibody specificity was tested by Western blot analysis.
2.7 Data analysis

2.7.1 Microarray analysis

Raw data from microarray analysis of 174 primary tumours using Affymetrix GeneChip Human X3P Arrays were analysed essentially as described (Farmer et al., 2009; Farmer et al., 2005). Briefly, raw data were processed with the statistical programming language R (www.r-project.org), and Bioconductor packages (www.bioconductor.org). Gene expression was normalised with the robust multi-array (RMA) method implemented in the package affy. Probe sets were filtered to eliminate redundancy based on annotation to defined Entrez-gene entries, and for each Entrez-gene only the most variable probe set was retained. Unsupervised clustering was performed using Cluster 3.0 for Mac and Java TreeView 1.1.3 with the following parameters: standard deviation greater than 1.5, median centre and normalize 3x, uncentred correlation, centroid linkage (de Hoon et al., 2004; Eisen et al., 1998; Saldanha, 2004).

Selected data from microarray analysis of 51 breast cancer cell lines were taken from (Neve et al., 2006) and processed in Cluster 3.0 and Java Treeview as described above.

Illumina microarray data of TOX3 expression in MCF-7 cells were filtered using the limma package in R. Normalised data were further interpreted using Cluster and Java Treeview. Alternatively, data were analysed by gene set enrichment analysis (GSEA) (Subramanian et al., 2005). Raw Illumina data (37,249 genes) were reduced to 7065 differentially expressed genes by excluding all genes with a standard deviation of smaller than 0.1. The resulting data set was used as input for GSEA. Duplicates were removed based on HUGO gene symbols, resulting in a final set of 5988 genes. This gene set was used for GSEA.

2.7.2 Multiple sequence alignment of amino acid and nucleotide sequences

In order to identify evolutionary conservation, multiple sequence alignments of amino acid and nucleotide sequences were carried out using ClustalW, Version 2.1 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Chenna et al., 2003).
2.7.3 Analysis of amino acid and nucleotide sequences

Plasmid sequences and chromatograms were analysed using the demo version 4.8 of Sequencher (Gene Codes Corporation, USA). Nucleotide and amino acid sequences were analysed using DNA Strider, version 1.4f6 (Douglas, 1995). Plasmid maps were created using A plasmid Editor, version 1.17 (ApE, http://biologylabs.utah.edu/jorgensen/wayned/ape/).

2.7.4 Analysis of protein properties

Physico-chemical properties of proteins and peptides were deduced from the protein or peptide sequence using the ProtParam tool on the ExPASy Proteomics Server (http://au.expasy.org/tools/protparam.html) (Gasteiger et al., 2005).

2.7.5 Prediction of alternative splice sites

Potential alternative splice sites in hUbiC-TOX3 and hUbiC-NTAP-TOX3 vectors were identified using two different internet-based splice site predictors (http://www.fruitfly.org/seq_tools/splice.html; http://spliceport.cs.umd.edu/) (Dogan et al., 2007; Reese et al., 1997).
Materials and Methods
3 Results

3.1 Expression of TOX3 in the normal breast and breast cancer

Several genome-wide association studies (GWAS) have identified TOX3 as a candidate breast cancer susceptibility gene. In addition, some of these studies suggest that the presence of SNPs in the regulatory region of the TOX3 gene is associated with an increased ER-positive breast cancer risk in particular. Using quantitative PCR analysis to compare TOX3 levels in human tissues, a recent study on TOX3 function in neurons showed that TOX3 is predominantly expressed in the brain (Dittmer et al., 2010). However, to date no information has been available on TOX3 gene expression in normal breast tissue or in breast cancer. Therefore, I was interested to investigate TOX3 gene expression patterns in normal as well as breast cancer cells.

3.1.1 TOX3 is expressed in the luminal cluster in ER-positive breast tumours

Microarray expression analysis is an important tool to identify new genes that exhibit a tumour subtype-specific expression pattern, or even define a tumour subtype. Microarray analysis of 162 primary breast tumours and 3 normal samples from the EORTC 10994 clinical trial shows that TOX3 clusters with classic luminal genes such as GATA3, FOXA1, TFF1 and ESR1, suggesting a shared function for TOX3 and these genes. Moreover, TOX3 expression appears to be restricted to luminal tumours (Fig. 11).
Figure 11. TOX3 is expressed in the luminal cluster in ER-positive tumours.

Expression microarray analysis of 162 primary breast tumours shows that TOX3 clusters with luminal genes. Tumour samples are a subset of pre-treatment biopsies from patients enrolled in the EORTC 10994 clinical trial. The three main tumour clusters from left to right are basal-like, molecular apocrine/HER2 and luminal.

3.1.2 Immunohistochemical analysis of tissue microarrays confirms TOX3 expression in luminal tumours

Immunohistochemical (IHC) staining of tumour samples for prognostic markers is a valuable technique in the clinical setting. IHC further provides information about the expression pattern of the protein of interest in situ in the tissue and cell type as well as about its subcellular localisation. I therefore wanted to develop an IHC staining protocol for the detection of TOX3 protein in tumour sections. I first tested a custom-made rabbit polyclonal peptide antibody TOX3 Rb57 as well as the only antibody that was commercially available at the time, the rabbit polyclonal peptide antibody ab77432 (Abcam), by IHC staining of HEK 293T cells transfected with the lentiviral TOX3 expression vector pXS-39, which is described in the following section (Fig. 12). The cells were formalin-fixed and paraffin-embedded prior to IHC staining. Neither antibody detected endogenous TOX3 protein in untransfected HEK 293T cells, while ectopically expressed TOX3 exhibited nuclear staining (Fig. 12).
HEK 293T cells were transiently transfected with the lentiviral TOX3 expression vector pXS-39. 48 hours post transfection, cells were pelleted, formalin-fixed, paraffin-embedded, and sectioned. Staining with TOX3 Rb57 (left) or TOX3 ab77432 (Abcam) (right) antibody shows nuclear localisation of ectopically expressed TOX3. Low and high magnification are shown.

Preliminary IHC staining of breast tumour samples using the TOX3 ab77432 antibody under optimised conditions showed only diffuse staining resembling the background staining of untransfected HEK 293T cells (data not shown). The diffuse staining was irrespective of the breast tumour subtype. I therefore decided to continue with the TOX3 Rb57 antibody for staining of tissue microarrays (TMAs) containing biopsies of ER-positive and ER-negative tumours as well as normal control tissue. Luminal tumours exhibited both nuclear and cytoplasmic TOX3 staining, while basal tumours were TOX3-negative. TOX3 expression was also observed in normal luminal epithelial cells and surprisingly, in myoepithelial cells. Moreover, stromal tissue, in particular lymphocytes, stain positive for TOX3 (Fig. 13).

The finding that TOX3 localised to the cytoplasm was somewhat surprising as TOX3 as a putative transcription factor would be expected to localise predominantly to the nucleus, as I observed for ectopically expressed TOX3, but it is not possible to conclude from the IHC results whether cytoplasmic staining is due to the quality of the antibody or to regulatory mechanisms that may sequester TOX3 in the cytoplasm. In order to rule out staining artefacts that are due to antibody background, I affinity-purified the custom-made TOX3 Rb57 antibody, as described in detail in chapter 4. While the purified antibody proved useful for Western blot analysis of TOX3 expression, it appeared to be too dilute as it gave no signal in immunohistochemical
staining of TOX3 even in the transfected positive control HEK 293T cells (data not shown). Unfortunately, I was not able to repeat the antibody purification due to time constraints.

![Image of TOX3 staining](image)

Figure 13. TOX3 is expressed in luminal but not in basal tumours.

TOX3 immunohistochemical staining of a tissue microarray of primary breast tumours. Luminal tumours exhibit both nuclear and cytoplasmic TOX3 expression (left column), while basal tumours are TOX3-negative (middle column). TOX3 expression is also found in normal ductal tissue (right column, top panel) and stromal tissue (right column, bottom tissue). Luminal and basal tissue are shown at low (top panels) and high (bottom panels) magnification.

3.1.3 A model system for endogenous TOX3 expression

The establishment of a robust model system was crucial for the characterisation of TOX3 function both at the molecular and cellular level. Established breast cancer cell lines are very helpful as an initial model for the study of new genes. Unlike other luminal genes such as GATA3 or FOXA1, TOX3 is not expressed in the classic luminal breast cancer cell lines MCF-7 and T-47D (Fig. 14, 15). Therefore I turned to the comprehensive 2006 study of 51 breast cancer cell lines by Gray and colleagues (Neve et al., 2006). Notably, all TOX3-overexpressing cell lines in their study were classified as luminal, with the exception of two cell lines (SUM225 and SUM190PT),
which were basal. Based on the expression data provided by Gray et al. (Fig. 14) and availability, I chose to further investigate TOX3 expression in the luminal breast cancer cell lines BT-474, ZR-75-1, MDA-MB-361, MDA-MB-415 and UACC-812.

![Figure 14. TOX3 expression in 51 breast cancer cell lines.](image)

Unsupervised hierarchical clustering of 51 breast cancer cell lines based on their expression of TOX3 and a selection of genes defining the luminal and molecular apocrine breast cancer subtypes. Data taken from (Neve et al., 2006).

Western blot analysis of a selection of breast cancer cell lines confirmed the absence of TOX3 expression in MCF-7, T-47D, CAMA-1, MCF-10A, MDA-MB-231, MDA-MB-468, and MDA-MB-453 cells (Fig. 15 A, Table 7). Consistent with the expression data in the study by Gray and co-workers, TOX3 expression was detectable in BT-474, MDA-MB-361, MDA-MB-415, ZR-75-1 and UACC-812 cells to a varying degree (Fig. 15 B). Relative TOX3 expression appeared to be highest in BT-474 and MDA-MB-361 cells, followed by ZR-75-1, MDA-MB-415 and UACC-812 cells.
Figure 15. Western blot analysis of TOX3 expression in breast cancer cell lines.

Western blot analysis of TOX3 protein expression in different breast cancer cell lines. (A) Western blot analysis of cell lines that do not express TOX3 protein. (B) TOX3-positive cell lines express TOX3 at varying levels. Whole cell lysates were prepared by direct lysis in 2x loading buffer. The unpurified TOX3 Rb57 antibody was used for the Western blot in (A), while affinity-purified antibody TOX Rb57 was used for Western blot analysis shown in (B). The open arrowhead marks a non-specific band and the filled arrowhead indicates a TOX3-specific band.

Based on TOX3 expression levels and growth properties, I chose the following breast cancer cell lines as model systems to investigate TOX3 function. MCF-7 cells were selected for the majority of TOX3 overexpression experiments, as they do not express TOX3 but may provide an adequate luminal environment. BT-474 and MDA-MB-361 were chosen as luminal cell model systems for endogenous TOX3 expression because they exhibited the highest relative TOX3 protein levels in Western blot analysis (Fig. 15). MDA-MB-231 cells were used as a TOX3-negative basal-like tumour cell model. Immortalised MCF-10A cells do not express TOX3 (Fig. 42 B) and were used for differentiation assays.
Table 7. TOX3 status of breast cell lines

<table>
<thead>
<tr>
<th>Cell line</th>
<th>TOX3</th>
<th>ER</th>
<th>PGR</th>
<th>HER2</th>
<th>Gene cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT-474</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Luminal</td>
</tr>
<tr>
<td>MDA-MB-361</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>Luminal</td>
</tr>
<tr>
<td>MDA-MB-415</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>Luminal</td>
</tr>
<tr>
<td>UACC-812</td>
<td>+/-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>Luminal</td>
</tr>
<tr>
<td>ZR-75-1</td>
<td>+/-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>Luminal</td>
</tr>
<tr>
<td>CAMA-1</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>Luminal</td>
</tr>
<tr>
<td>MDA-MB-231</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Basal</td>
</tr>
<tr>
<td>MDA-MB-453</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Luminal</td>
</tr>
<tr>
<td>MDA-MB-468</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Luminal</td>
</tr>
<tr>
<td>MCF-7</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>Luminal</td>
</tr>
<tr>
<td>MCF-10A</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>Luminal</td>
</tr>
<tr>
<td>T-47D</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Luminal</td>
</tr>
</tbody>
</table>

Data are taken from (Neve et al., 2006). TOX3 expression data were confirmed by Western blot analysis.

3.1.4 TOX3 expression in the normal breast epithelium

In addition to investigating TOX3 function in breast cancer, I aimed to understand the role of TOX3 in the normal breast epithelium. Using Western blot analysis I was not able to detect TOX3 expression in unsorted normal human mammary epithelial cells maintained in conventional HMM+ medium (termed HMEC in the following) or WIT medium (BPEC) (Fig.16 A) (Duss et al., 2007; Ince et al., 2007). Quantitative PCR analysis of unsorted cells freshly isolated from reduction mammoplasty tissue confirmed this finding (data not shown). In contrast, preliminary analysis of FACS-sorted mammary epithelial cells using the surface markers CD10 and EpCAM to separate stem and progenitor-enriched CD10+ myoepithelial/basal and mature luminal EpCAM+ subpopulations, respectively, suggested that relative TOX3 expression was highest in the mature luminal subset (data not shown). Owing to the unavailability of sufficient mammoplasty sample, the data could not be normalised.

However, the observed trend was consistent with expression array data from the Smalley group, which showed that relative TOX3 expression was highest in luminal ER+ cells (CD24$^{\text{high}}$ Sca-1$^+$), while luminal ER- cells (CD24$^{\text{high}}$ Sca-1$^-$) show significantly lower TOX3 levels and myoepithelial cells (CD24$^{\text{low}}$ Sca-1$^-$) do not express TOX3 (Fig. 16 B) (Kendrick et al., 2008). A recent study by the Visvader group further confirmed that TOX3 expression was upregulated in the mature luminal
compartment (CD49f EpCAM^+^) but not in the MaSC-enriched (CD49f^^{\text{high}} EpCAM^+) and luminal progenitor (CD49f^+ EpCAM^+) subpopulations (Fig. 16 C). TOX3 did not appear to be expressed in stromal cells (CD49f^− EpCAM^−^) (Fig. 16 C). Importantly, TOX3 was the 7th most conserved gene between the human and mouse mature luminal subsets (Table 8) (Lim et al., 2010). Interestingly, CITED1, which has been shown to cooperate with TOX3 to activate transcription in neurons (Dittmer et al., 2008) ranked 10th in the list of conserved luminal genes (Table 8). The expression data from the Visvader group further showed that expression of LOC643714, the putative non-coding RNA gene just upstream of the TOX3 gene, did not change across epithelial subpopulations (Fig. 16 C).
Figure 16. TOX3 expression in the normal human and mouse mammary gland

(A) Western blot analysis shows that unsorted BPEC and HMEC do not express endogenous TOX3 at levels that are detectable with the unpurified TOX3 Rb57 antibody. BPEC stably expressing TOX3 under the control of the human PGK promoter (pXS-36) were used as a positive control. (B) Microarray analysis of mouse mammary epithelial subpopulations isolated using CD24 and Sca1, as described in the text. Relative expression of selected luminal genes and TOX3 in mature epithelial subpopulations is shown. Expression data were taken from (Kendrick et al., 2008). (C) Transcriptional profiling of human mammary epithelial subpopulations sorted using the markers CD49f and EpCAM, as described the text. Relative expression of selected genes including TOX3 is shown. Expression data were taken from (Lim et al., 2010).
### Table 8. Conserved genes in the mature luminal epithelial subpopulation

<table>
<thead>
<tr>
<th>Symbol Human</th>
<th>Log Fold Change</th>
<th>Symbol Mouse</th>
<th>Log Fold Change</th>
<th>Average Log Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOXA1</td>
<td>4.85</td>
<td>Foxa1</td>
<td>2.90</td>
<td>3.87</td>
</tr>
<tr>
<td>DNAJC12</td>
<td>4.44</td>
<td>Dnajc12</td>
<td>2.57</td>
<td>3.51</td>
</tr>
<tr>
<td>MLPH</td>
<td>5.30</td>
<td>Mlph</td>
<td>1.67</td>
<td>3.48</td>
</tr>
<tr>
<td>SPINK1</td>
<td>4.69</td>
<td>Spink3</td>
<td>2.05</td>
<td>3.37</td>
</tr>
<tr>
<td>RASEF</td>
<td>4.55</td>
<td>Rasef</td>
<td>2.14</td>
<td>3.34</td>
</tr>
<tr>
<td>BATF</td>
<td>4.84</td>
<td>Batf</td>
<td>1.45</td>
<td>3.15</td>
</tr>
<tr>
<td>TOX3</td>
<td>4.09</td>
<td>Tox3</td>
<td>2.13</td>
<td>3.11</td>
</tr>
<tr>
<td>HMGCS2</td>
<td>3.51</td>
<td>Hmgcs2</td>
<td>2.58</td>
<td>3.05</td>
</tr>
<tr>
<td>EEF1A2</td>
<td>5.13</td>
<td>Eef1a2</td>
<td>0.85</td>
<td>2.99</td>
</tr>
<tr>
<td>CITED1</td>
<td>2.90</td>
<td>Cited1</td>
<td>3.05</td>
<td>2.98</td>
</tr>
<tr>
<td>ABCC8</td>
<td>3.59</td>
<td>Abcc8</td>
<td>2.23</td>
<td>2.91</td>
</tr>
<tr>
<td>PRLR</td>
<td>3.22</td>
<td>Prlr</td>
<td>2.59</td>
<td>2.90</td>
</tr>
<tr>
<td>SLC7A2</td>
<td>3.65</td>
<td>Slc7a2</td>
<td>2.07</td>
<td>2.86</td>
</tr>
<tr>
<td>SLC16A5</td>
<td>2.57</td>
<td>Slc16a5</td>
<td>2.87</td>
<td>2.72</td>
</tr>
<tr>
<td>ALDH3B2</td>
<td>4.01</td>
<td>Aldh3b2</td>
<td>1.18</td>
<td>2.59</td>
</tr>
<tr>
<td>SLC40A1</td>
<td>3.58</td>
<td>Slc40a1</td>
<td>1.42</td>
<td>2.50</td>
</tr>
<tr>
<td>WNT4</td>
<td>2.20</td>
<td>Wnt4</td>
<td>2.80</td>
<td>2.50</td>
</tr>
<tr>
<td>SPDEF</td>
<td>2.79</td>
<td>Spdef</td>
<td>2.17</td>
<td>2.48</td>
</tr>
<tr>
<td>REEP6</td>
<td>2.98</td>
<td>Reep6</td>
<td>1.95</td>
<td>2.46</td>
</tr>
<tr>
<td>SULT2B1</td>
<td>2.90</td>
<td>Sult2b1</td>
<td>1.94</td>
<td>2.42</td>
</tr>
<tr>
<td>PGR</td>
<td>2.22</td>
<td>Pgr</td>
<td>2.30</td>
<td>2.26</td>
</tr>
<tr>
<td>PVALB</td>
<td>1.59</td>
<td>Pvalb</td>
<td>2.89</td>
<td>2.24</td>
</tr>
<tr>
<td>MBOAT1</td>
<td>2.03</td>
<td>Mboat1</td>
<td>2.33</td>
<td>2.18</td>
</tr>
<tr>
<td>TNFSF11</td>
<td>1.25</td>
<td>Tnfsf11</td>
<td>3.07</td>
<td>2.16</td>
</tr>
<tr>
<td>TGM2</td>
<td>3.10</td>
<td>Tgm2</td>
<td>1.21</td>
<td>2.15</td>
</tr>
<tr>
<td>FLVCR2</td>
<td>2.85</td>
<td>Mfsd7c</td>
<td>1.41</td>
<td>2.13</td>
</tr>
<tr>
<td>FAAH</td>
<td>1.51</td>
<td>Faah</td>
<td>2.73</td>
<td>2.12</td>
</tr>
<tr>
<td>SLC22A18</td>
<td>1.96</td>
<td>Slc22a18</td>
<td>2.17</td>
<td>2.07</td>
</tr>
<tr>
<td>ESR1</td>
<td>2.39</td>
<td>Esr1</td>
<td>1.69</td>
<td>2.04</td>
</tr>
<tr>
<td>MYB</td>
<td>2.44</td>
<td>Myb</td>
<td>1.59</td>
<td>2.01</td>
</tr>
</tbody>
</table>

Table is adapted from and is described in (Lim et al., 2010). Fold changes were determined by pairwise comparison of gene expression in epithelial subpopulations for each species.
3.2 Identification of TOX3 interacting proteins by tandem affinity purification

The sequencing of the human genome has made it possible to analyse complex biological networks of genes and the proteins they encode on a genome-wide basis. The identification of interactions between proteins is one way to obtain insight into their function. Traditionally, protein-protein interactions and new interacting partners of a given protein have been identified using large-scale two-hybrid screening (Ito et al., 2000; Uetz et al., 2000). Limitations of this method include the requirement of large libraries of fusion constructs as well the low degree of information provided, for instance the lack of information about the stoichiometry of protein interactions. Biochemical purification in combination with mass spectrometric analysis allows for the identification of interacting partners of a given protein in a proteome-wide approach. The availability of complete genomic and proteomic sequences facilitates rapid comparison of mass spectrometric profiles of the protein complexes and the identification of the contained proteins. The limiting factor in this case is the protein purification step, which requires the knowledge of the biochemical properties of target proteins. Affinity purification of proteins using an affinity tag largely circumvents this problem.

To gain clues to the function of TOX3, I therefore chose tandem affinity purification (TAP) in combination with mass spectrometric analysis to purify TOX3 and identify its interacting proteins. TAP was originally developed in yeast for the purification of protein complexes under native conditions, even when expressed at physiological levels, without the requirement of prior knowledge of complex composition or function (Puig et al., 2001; Rigaut et al., 1999). Tandem affinity purification involves a two-step affinity purification of the protein of interest that is fused to a bipartite affinity tag. In the original study, the TAP tag consisted of two IgG-binding units of *Staphylococcus aureus* protein A and calmodulin binding peptide (CBP), which were separated by a TEV protease cleavage site, allowing for the purification of the protein of interest using an IgG column, followed by TEV protease cleavage of the protein A tag and subsequent purification using a calmodulin column (Rigaut et al., 1999). I used the InterPlay Mammalian TAP protocol (Agilent Technologies), a modified tandem affinity purification protocol involving tandem purification via the affinity tags
streptavidin binding peptide (SBP) and CBP. SBP is a synthetic peptide that was identified in a screen for peptides that bind to streptavidin with high affinity ($2 \times 10^{-9}$ M) and can be eluted from streptavidin with biotin (Keefe et al., 2001; Wilson et al., 2001). The CBP tag is derived from a C-terminal fragment of myosin light-chain kinase and binds to calmodulin with high affinity ($1 \times 10^{-9}$ M) in the presence of calcium (Stofko-Hahn et al., 1992). Elution from the calmodulin resin is achieved by chelation of calcium with EGTA. I decided to use the InterPlay TAP protocol because it was optimised for the purification of proteins expressed at physiological levels in mammalian cells. In addition, since biochemical properties of TOX3 protein were unknown, I considered gentle wash conditions and small molecule elution instead of protease cleavage as advantageous. The TAP strategy is depicted in Figure 17.

![Tandem affinity purification protocol](image.png)

**Figure 17.** Tandem affinity purification protocol.

Schematic of the tandem affinity purification protocol. NTAP-TOX3 is expressed in mammalian cells. In the first purification step, whole cell extracts are incubated with streptavidin resin. Following three washes, NTAP-TOX3 and interacting protein complexes are eluted using biotin. In the second purification step, NTAP-TOX3 is allowed to bind to calmodulin resin and subsequently eluted using EGTA. Eluted protein is concentrated, separated by SDS-PAGE and analysed by mass spectrometry.
3.2.1 Lentiviral NTAP expression constructs

The goal of the tandem affinity purification of TOX3 was to identify the proteins in complex with TOX3 under physiological conditions. I ultimately intended to carry out the TAP purification using breast cancer cells as well as normal mammary epithelial cells. Therefore, I chose to express NTAP-TOX3 using lentiviral vectors, which readily infect dividing and quiescent cells and allow for the efficient creation of cell lines stably expressing the gene of interest. I created three sets of lentiviral expression vectors coding for TOX3 fused to an N-terminal TAP tag and their respective control vectors, which coded for the TAP tag only (Fig. 18 A and Appendix). The different vectors allowed the expression of NTAP-TOX3 at high levels using the CMV (pXS-85 and pXS-72 as control) or human UbiC promoters (pXS-96 and pXS-99), or at low levels using the human PGK promoter (pXS-95 and control pXS-98). The human PGK promoter had previously been shown to be active in mammary epithelial progenitor and mature cells (Duss et al., 2007). Since the TAP tag (Fig. 18 B) was cloned from the parental pNTAP vector contained in the Mammalian Interplay TAP system (Agilent Technologies) only minimal optimisation of the purification protocol was expected.

Transient transfection of the different TAP constructs into HEK 293T cells and subsequent Western blot analysis confirmed that both TOX3 and the TAP tag were expressed from all constructs (Fig. 19). The hUbiC and CMV promoters drove expression at a level that was about 10-fold higher than the level of hPGK promoter-controlled expression. NTAP-TOX3 expression from both the CMV and the hPGK vector resulted in a band of approximately 75 kDa, consistent with the predicted size of 72 kDa (Fig. 18 C). In addition, CMV-NTAP-TOX3 produced two smaller bands at approximately 50 and 60 kDa (Fig. 19), which may be due to degradation products. Surprisingly, expression of NTAP-TOX3 from the hUbiC vector resulted in several bands, the biggest of which with an apparent molecular weight of approximately 100 kDa. Like the CMV construct, the hUbiC-NTAP-TOX3 vectors produced several smaller bands, the most prominent one of which ran at approximately 80 kDa. Western blot analysis of TAP tag expression using an antibody against the CBP moiety of the tag confirmed this banding pattern (Fig. 19). Western blot analysis of the control constructs using the CBP antibody detected a band of approximately 35
kDa for the hUbiC and CMV control constructs, while the TAP tag was not detected by the CBP antibody in the hPGK control lane (Fig. 19).

Figure 18. NTAP-TOX3 lentiviral expression vectors.

(A) Schematic of the three different lentiviral NTAP-TOX3 expression constructs (pXS-85, pXS-95 and pXS-96) used in this study; the hPGK-NTAP control vector (pXS-98) is shown as an example of the corresponding control vectors. For an enlarged version of the maps, see Appendix. (B) Peptide sequence of the TAP tag components streptavidin binding peptide (SBP), calmodulin binding peptide (CBP) and linker regions. (C) Schematic of the NTAP-TOX3 protein domains. Molecular weights were determined using the ProtParam tool. HIV, Human immune deficiency virus; SV40, Simian vacuolating virus 40; ORI, origin of replication; ampR, ampicillin resistance gene; RSV, Rous sarcoma virus; LTR, long terminal repeat; SIN LTR, self-inactivating LTR; psi, packaging element; RRE, Rev responsive element; cPPT, central polypurine tract; attB, Gateway recombination site; WPRE, Woodchuck hepatitis post-transcriptional regulatory element; mPGK prom, murine phosphoglycerate kinase 1 promoter; hPGK prom, human PGK promoter; CMV prom, cytomegalovirus promoter; hUbiC, human ubiquitin C promoter; puroR (pac), puromycin resistance gene (puromycin N-acetyltransferase).
Since the correct plasmid sequence had been confirmed by DNA sequencing, the fact that the band corresponding to the TAP tag expressed from the hPGK vector was not detected was presumably due to the comparatively low expression level of the hPGK construct and the poor antibody quality.

Western blot analysis of TOX3 and CBP expression of the different lentiviral NTAP-TOX3 and control constructs. Western blot analysis for CBP did not detect the TAP tag in the hPGK control lane.

Western blot analysis for ubiquitin ruled out posttranslational modification by ubiquitin as an explanation for the high molecular weight isoforms seen for the hUbiC-NTAP-TOX3 construct (data not shown). Notably, both the hUbiC and the CMV lentiviral vector backbone also contained the marker gene GFP under the control of the human PGK promoter (Fig. 18 A), which was expected to result in green fluorescence of cytoplasmic GFP upon transfection. HEK 293T cells transfected with the hUbiC-NTAP control vector exhibited cytoplasmic GFP fluorescence, while GFP in the hUbiC-NTAP-TOX3 transfected cells localised almost entirely to the nucleus (Fig. 20 A). This led me to speculate that this was due to an in-frame fusion of the TOX3 ORF with the downstream GFP ORF through alternative splicing, which resulted in a functional TOX3-GFP fusion protein that localised to the nucleus due to the presence of the nuclear localisation signal (NLS) in TOX3.

![Western blot analysis of NTAP constructs.](image)
other TOX family members have been shown to localise to the nucleus (Kajitani et al., 2004; Lee et al., 2009; Wilkinson et al., 2002).

Figure 20. The hUbiC lentiviral vector produces a TOX3-GFP fusion protein.

(A) HEK 293T cells were transiently transfected with the hUbiC-NTAP control vector (left) or hUbiC-NTAP-TOX3. Subcellular localisation of GFP was analysed using an Olympus CKX41 fluorescent microscope at 20 × magnification (top panel, phase contrast; bottom panel, GFP). (B) Western blot analysis of GFP expressed from hUbiC-TAP constructs. Whole cell lysates from transiently transfected HEK 293T cells were separated by 10% SDS-PAGE analysed by Western blot. (C) Western blot analysis of TOX3 transiently expressed in HEK 293T cells from the hUbiC-TOX3 vector (pXS-39) shows several isoforms of TOX3. (D) Confocal microscopy analysis of GFP subcellular localisation in MCF-7 cells transiently transfected with the hUbiC-TOX3 (pXS-39, bottom panel) or a hUbiC-DsRed2 control vector (pXS-41, top panel). (E) Prediction of splice sites in the lentiviral plasmid sequence. The 3′ end of the TOX3 ORF is shown in yellow, the 5′ end of the GFP ORF is highlighted in green. Predicted splice donor/acceptor sites are marked in red.
Analysis of the lentiviral plasmid sequence using splice site prediction software indeed revealed the presence of putative splice donor and acceptor sites at the 3’ end of the TOX3 ORF and just upstream of the 5’ end of the GFP ORF, respectively (Fig. 20 E). Alternative splicing at these sites would delete only part of the last exon and stop codon of the TOX3 ORF. The presence of a TOX3-GFP fusion protein upon transfection of 293T cells was confirmed by Western blot analysis for GFP (Fig. 20 B). Using the ProtParam tool, the size of the NTAP-TOX3-GFP fusion construct is predicted to be 99.2 kDa, which is consistent with the highest running band in the Western blot (Fig. 20 B). Correctly spliced GFP was expected to run at 27 kDa, which could be seen both in the control and NTAP-TOX3 lanes. Alternative splicing of the transcript prior to integration of the viral vector would also explain the observation that functional hUbiC-TOX3 lentiviral particles could, if at all, only be produced at very low titres as measured by green fluorescence of infected cells. Alternative splicing appeared to also affect the hUbiC-NTAP control vector, as part of the TAP tag, appeared to be fused to GFP and gave rise to a band at approximately 35 kDa in the Western blot (Fig. 19, 20).
Due to the GPF fusion, the hUbiC-NTAP-TOX3 construct was not used for the final tandem affinity purification experiments but prior to finding an explanation for the high molecular weight bands observed in the Western blot, I used the construct for preliminary optimisation of the TAP conditions because of its high protein yield.

As discussed in the following chapter, I also created a hUbiC-TOX3 lentiviral construct (pXS-39) lacking the TAP tag (Fig. 31). Since the hUbiC-TOX3 vector differed from the hUbiC-NTAP-TOX3 vector only in the absence of the TAP tag, it was not surprising that the hUbiC-TOX3 vector, too, was alternatively spliced and produced a TOX3-GFP fusion protein. Western blot analysis for TOX3 confirmed the presence of the corresponding high molecular weight bands (Fig. 20 C), and microscopic analysis of GFP localisation in MCF-7 cells transiently transfected with the hUbiC-TOX3 vector showed exclusive nuclear localisation of GFP (Fig. 20 D). In contrast, GFP expressed from the corresponding control vector (pXS-41) localised both to the cytoplasm and the nucleus (Fig. 20 D).

Figure 21. The CMV-NTAP control lentiviral vector produces an NTAP-GFP fusion protein.

Western blot analysis of purified NTAP-TOX3 and control NTAP tag, which had been expressed from the CMV-NTAP vector set in HEK 293T cells. A prominent band at approximately 35 kDa in the control lane is detected by both the CBP and GFP antibodies and likely represents a TAP-tag GFP fusion protein.

As the CMV-NTAP control vector shared large parts of the backbone of the hUbiC vector including the GFP ORF downstream of the TAP tag ORF (Fig. 19 A), it was not surprising that it gave rise to a GFP fusion protein (Fig. 21). As a consequence, the set of CMV-TAP vectors was not used for further TAP experiments. Curiously,
the alternative splice did not seem to take place in the CMV-NTAP-TOX3 vector (Fig. 21).

3.2.2 Optimisation of TAP purification conditions

For all initial TAP experiments that were carried out to optimise the purification conditions, HEK 293T cells were calcium phosphate transfected with the respective NTAP-TOX3 and control constructs, and cell lysates were prepared 48 hours post transfection. Pilot TAP purification experiments demonstrated that NTAP-TOX3 could be successfully purified using the standard conditions recommended for the Agilent InterPlay Mammalian TAP System. The first purification step using streptavidin resin removed a large part of contaminants. Further purification was achieved by binding of TAP-tagged TOX3 to the calmodulin resin (Fig. 22).

However, initial experiments also demonstrated that three steps in the purification protocol limited the final protein yield. The lysis buffer supplied with the InterPlay Mammalian TAP purification kit contained 150 mM NaCl in addition to 10 mM Tris-HCl, pH 8.0, 0.1% NP-40 and 2 mM EDTA. It was not possible to extract the majority of TOX3 protein from the nucleus under these conditions, as a significant part of the protein appeared to remain in the cell debris pellet after centrifugation of the cell lysates (Fig. 22). I hypothesised that this was due to a strong association of TOX3 with chromatin. In addition, elution from both streptavidin and calmodulin resins was not efficient. This is a known caveat. Recovery rates of 50% for the CBP-tag, for instance, have been reported (Rigaut et al., 1999). However, the recovery rate in my pilot experiments after binding to the calmodulin resin was below 50% as judged by silver stain and Western blot analysis. Therefore, I tested several modifications of the standard protocol but at the same time, I avoided to change the purification conditions too drastically so as not to disrupt potential TOX3-containing protein complexes.
Figure 22. Tandem affinity purification pilot experiment.

Silver stain analysis of a typical TAP purification. HEK 293T cells were transfected with the UbiC-NTAP-TOX3 (pXS-96) construct. Tandem affinity purification was carried out using standard conditions recommended for the InterPlay Mammalian TAP system. Individual purification steps were verified by silver staining following 10% SDS-PAGE. Amounts given in the table are relative to the total amount of sample at the respective purification step (equalling the starting material minus loss occurring at each step). The asterisk indicates the NTAP-TOX3-GFP fusion protein. Arrows indicate putative isoforms of NTAP-TOX3. Negative control samples are not shown.

To optimise cell lysis and to increase solubility of TOX3, I compared the following lysis conditions: 150 mM NaCl/freeze-thaw (standard protocol), 400 mM NaCl/freeze-thaw, 400 mM KCl (modified Dignam and Roeder B/C buffer for the preparation of nuclear extracts)/freeze-thaw and NET buffer/freeze-thaw (Dignam et al., 1983). All conditions were tested without (standard protocol) or with sonication of the cell lysate (Fig. 23 A). Following lysis, cell lysates were centrifuged and both the pellet, which was expected to contain cell debris and genomic DNA and the non-dissociated chromatin fractions, and the supernatant containing soluble protein were subjected to Western blot analysis for TOX3. I expected to see higher relative amounts of TOX3 in the supernatant in contrast to lower relative amounts in the pellet.
Results

as a result of more efficient extraction of TOX3 from cellular chromatin. However, no significant difference was observed between the different lysis conditions (Fig. 23 A, results for NET buffer not shown). Therefore, the original TAP lysis buffer was used for further large-scale TAP purification experiments. The salt concentration was increased to 400 mM NaCl nevertheless as some preliminary experiment had suggested that extraction of TOX3 from cellular chromatin was more efficient using this condition.

Figure 23. Optimisation of TAP lysis and elution conditions.

(A) TOX3 Western blot analysis of HEK 293T cells transiently transfected with the hPGK-NTAP-TOX3 (top) or hUbiC-NTAP-TOX3 vector (bottom). Cells were lysed using different salt concentrations with or without sonication, as described in the text. (B) TOX3 Western blot analysis of HEK 293T cells transiently transfected with the hUbiC-NTAP-TOX3 vector. Elution from the streptavidin resin was tested using different salt and biotin concentrations. Equal amounts were loaded across sample groups (for instance all eluate fractions) but not between samples from different purification steps, which are thus not directly comparable.

To optimise the streptavidin elution efficiency, streptavidin elution buffer (SEB) was supplemented with 150 mM NaCl (standard concentration) or 400 mM NaCl. Alternatively, standard SEB was supplemented with 4 mM instead of 2 mM biotin, alternatively (Fig. 23 B). Doubling the biotin concentration did not increase the TOX3 protein yield in the eluate, which is probably explained by that fact that the high affinity binding of biotin to streptavidin (dissociation constant $K_d = 10^{-15}$ M) is already sufficient to replace SBP at all binding sites at a biotin concentration of 2 mM (Weber et al., 1989). Relative amounts of TOX3 protein remaining bound to the
streptavidin resin after the elution appeared to be lower in the presence of 400 mM NaCl but the relative amount of TOX3 was not higher in the corresponding eluate (Fig. 23 A). Due to the inconclusive results, the standard salt concentration of 150 mM NaCl was used for large-scale experiments. To block non-specific binding of chromatin to the streptavidin beads, thereby reducing the binding capacity of the resin for NTAP-TOX3, cell lysates were supplemented with 100 µg/ml salmon sperm DNA during incubation with the streptavidin beads. This appeared to slightly increase the yield of TOX3 in the streptavidin eluate (data not shown) but the difference was deemed negligible and salmon sperm DNA was not used in later TAP purifications.

Increasing the concentration of chelating agent EGTA, which was used to disrupt calcium-dependent binding of CBP to the calmodulin resin, was two-fold to 10 mM appeared to augment the TOX3 protein yield in the calmodulin eluate fraction (data not shown). The calmodulin elution buffer (CEB) was modified accordingly in further TAP purifications.

3.2.3 Tandem affinity purification of TOX3

To identify proteins that interact with TOX3, a large-scale TAP purification of NTAP-TOX3 was carried out. HEK 293T cells were transiently transfected with the hPGK-NTAP-TOX3 or hPGK-NTAP-TOX3 lentiviral vector. Transfected HEK 293T cells were chosen for the initial mass spectrometry analysis because HEK 293T cells can be easily produced in large quantities, and they can be transfected using calcium phosphate at an efficiency nearing 100%, thus maximising the final TOX3 protein yield. 48 hours following transfection, approximately $2 \times 10^8$ cells per condition were harvested using the optimised conditions. Correct NTAP-TOX3 expression was confirmed by silver stain and Western blot analysis (Fig. 24).
Figure 24. Silver stain and Western blot analysis of NTAP-TOX3 TAP.

TAP purification of TOX3 expressed from the hPGK-NTAP-TOX3 lentiviral vector in HEK 293T cells. Proteins were separated by 4-12% gradient gel electrophoresis. Proteins present after each of the single affinity purification steps as well as concentrated purified protein are visualised by silver staining. Western blot analysis for CBP does not show the TAP tag in the control lane due to its small size of approximately 7 kDa.

3.2.4 Mass spectrometric analysis of purified TOX3 and interacting proteins

For mass spectrometric analysis, TOX3 protein purified from $10^8$ cells was separated by gradient gel electrophoresis, stained with Coomassie Brilliant Blue G-250 (colloidal blue) and analysed by nano-scale liquid chromatographic tandem mass spectrometry (nLC-MS/MS). For a detailed description, see the Materials and Methods and Appendix chapters. Identified peptide sequences were compared to the UniProtKB/Swiss-Prot human protein database.
Results

Table 9. Mass spectrometry analysis of proteins associated with NTAP-TOX3

<table>
<thead>
<tr>
<th>Gel section</th>
<th>Official gene symbol</th>
<th>Name</th>
<th>Accession no.</th>
<th>Mol mass (Da)</th>
<th>No. peptides matched</th>
<th>No. different peptides</th>
<th>% Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-7, 16</td>
<td>CALM1</td>
<td>Calmodulin</td>
<td>P62158</td>
<td>16,826</td>
<td>64</td>
<td>9</td>
<td>60</td>
</tr>
<tr>
<td>3, 7-16</td>
<td>HSPA8</td>
<td>Heat shock 70kDa protein B</td>
<td>P11142</td>
<td>70,854</td>
<td>126</td>
<td>28</td>
<td>48</td>
</tr>
<tr>
<td>6, 7</td>
<td>RPS3</td>
<td>Ribosomal protein S3</td>
<td>A7E2S3</td>
<td>26,671</td>
<td>16</td>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td>6-16</td>
<td>HSPA1A</td>
<td>Heat shock 70kDa protein 1A</td>
<td>P08107</td>
<td>70,009</td>
<td>105</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>8-10</td>
<td>YBX1</td>
<td>Y box-binding protein I</td>
<td>P67809</td>
<td>35,903</td>
<td>24</td>
<td>9</td>
<td>43</td>
</tr>
<tr>
<td>3, 8-16</td>
<td>HSPA5</td>
<td>Heat shock 70kDa protein 5</td>
<td>Q24JP5</td>
<td>72,289</td>
<td>54</td>
<td>22</td>
<td>40</td>
</tr>
<tr>
<td>3, 5, 8, 12-14</td>
<td>UBB</td>
<td>Ubiquitin B</td>
<td>P0CG47</td>
<td>25,746</td>
<td>8</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>4-6, 11, 12</td>
<td>HSPA9</td>
<td>heat shock 70kDa protein 9</td>
<td>P38646</td>
<td>73,635</td>
<td>27</td>
<td>20</td>
<td>37</td>
</tr>
<tr>
<td>3, 5, 8, 12-14</td>
<td>UBC</td>
<td>Ubiquitin C</td>
<td>P0CG48</td>
<td>76,982</td>
<td>8</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>HSBP1</td>
<td>Heat shock factor binding protein 1</td>
<td>O75506</td>
<td>22,768</td>
<td>4</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>8-10, 13-16</td>
<td>ACTB</td>
<td>Actin, beta</td>
<td>P60709</td>
<td>41,710</td>
<td>46</td>
<td>13</td>
<td>36</td>
</tr>
<tr>
<td>8-10, 13-16</td>
<td>ACTG1</td>
<td>Actin, gamma 1</td>
<td>P63261</td>
<td>41,766</td>
<td>46</td>
<td>13</td>
<td>35</td>
</tr>
<tr>
<td>7, 8</td>
<td>RPSA</td>
<td>Ribosomal protein SA</td>
<td>P08865</td>
<td>32,833</td>
<td>11</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>8-10, 13, 14</td>
<td>TUBA1C</td>
<td>Tubulin, alpha 1c</td>
<td>Q9BQE3</td>
<td>49,864</td>
<td>34</td>
<td>13</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>RPS18</td>
<td>Ribosomal protein S18</td>
<td>P62269</td>
<td>17,708</td>
<td>10</td>
<td>6</td>
<td>33</td>
</tr>
<tr>
<td>7, 8, 10</td>
<td>HNRNPA1</td>
<td>Heterogeneous nuclear ribonucleoprotein A1</td>
<td>P09651</td>
<td>29,368</td>
<td>16</td>
<td>8</td>
<td>31</td>
</tr>
<tr>
<td>2-16</td>
<td>TOX3</td>
<td>TOX high mobility group box family member 3</td>
<td>Q15405</td>
<td>72,054</td>
<td>416</td>
<td>15</td>
<td>30</td>
</tr>
<tr>
<td>3, 5, 8, 12-14</td>
<td>UBA52</td>
<td>Ubiquitin A-52 residue ribosomal protein fusion product 1</td>
<td>P62987</td>
<td>14,719</td>
<td>8</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>6-8</td>
<td>PHB2</td>
<td>Prohibitin 2</td>
<td>Q99623</td>
<td>33,276</td>
<td>13</td>
<td>7</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>ELAVL1</td>
<td>ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R)</td>
<td>Q15717</td>
<td>72,054</td>
<td>6</td>
<td>6</td>
<td>28</td>
</tr>
</tbody>
</table>

Results were sorted consecutively by coverage, number of peptides matched, and number of different peptides matched. The gel lane had been separated into 16 sections (numbered from bottom to top). Keratins were omitted from the list. Protein accession numbers refer to the UniProtKB database (release 2011_06).

Mass spectrometric analysis of the NTAP-TOX3 lane showed significant contamination with keratins (see Appendix for full list of MS results), which was reflected both in the number of scanned peptides and the percentage of coverage. The
Results

presence of a large amount of keratins, which is a common problem when using sensitive protein detection techniques such as mass spectrometry, likely masked the scans of true interacting proteins present in the sample and thus greatly complicated the interpretation of the data. Due to the significant contamination with keratins, the control lane was not analysed by mass spectrometry.

MS analysis of the NTAP-TOX3 lane further predominantly identified proteins that are expressed in abundance in the cell including heat shock and other chaperone proteins, proteosome components, cytoskeletal proteins such as actins, ribosomal proteins as well as translation factors (Table 9). Indeed, the only protein that was identified with what was considered sufficient confidence (coverage 43%, 9 unique peptide matches) and which did not belong to one of the latter groups of proteins was the transcription/translation factor Y box-binding protein 1 (YBX1, YB-1). YBX1 was considered an interesting candidate for a TOX3 interacting protein as several studies have implicated YBX1 in malignant transformation, with evidence for it being involved in epithelial-mesenchymal transition (EMT), mammary stem cell self-renewal and drug resistance (Evdokimova et al., 2009; To et al., 2010). Based on the transcriptional profiling data from the Visvader lab, YBX1 does not appear to exhibit a mammary subpopulation-dependent expression pattern (Fig. 16 C).

![Figure 25](image)

Figure 25. Transient expression of NTAP-TOX3 induces heat shock proteins.

Overexpression of NTAP-TOX3 following transient transfection of HEK 293T cells leads to the expression of heat shock proteins and subsequent enrichment of the chaperone proteins during TAP. An antibody detecting both HSPA1A (HSP70) and HSPA1B (HSP72) was used for Western blot analysis.

Western blot analysis confirms the enrichment of heat shock protein 70/72 (HSPA1A/B) in the TAP fractions (Fig. 25) but I did not pursue this result, as
Results

induction and co-purification of chaperone proteins is a known artefact when overexpressing tagged proteins for affinity purification, presumably because overexpression of the protein of interest also increases the likelihood of protein misfolding (Gingras et al., 2005; Puig et al., 2001).

Another caveat associated with the overexpression of the TAP-tagged protein by transient transfection are the resulting non-stoichiometric binding conditions, which may skew the association with interacting proteins and thus complicate the identification of physiological binding partners. To avoid the problem associated with overexpression of ectopic protein, I had originally designed the NTAP-TOX3 constructs as lentiviral vectors to create cell lines stably expressing TAP-TOX3. I am currently performing tandem affinity purification of TOX3 from HEK 293T and MCF-7 lines stably expressing NTAP-TOX3. As mentioned above, $10^8$ transfected HEK 293T cells had been used to purify protein for mass spectrometric analysis. Preliminary experiments showed that the same number of HEK 293T cells stably expressing NTAP-TOX3 under the control of the hPGK promoter did not produce protein amounts sufficient for mass spectrometric analysis (data not shown). Therefore, the goal of ongoing work is to obtain an adequate protein yield by increasing the number of cells stably expressing NTAP-TOX3 while at the same time ensuring an expression level of TOX3 that will not provoke protein misfolding and the association of chaperone proteins.

3.2.5 Analysis of TOX3 protein interaction partners

To confirm whether TOX3 interacted with the transcription/translation factor YBX1, I aimed to perform co-immunoprecipitation of the two proteins. In an initial experiment, I transiently expressed TOX3 in HEK 293T cells from the hPGK lentiviral vector. Immunoprecipitation of TOX3 using the custom-made rabbit polyclonal TOX3 Rb57 antibody failed, which was most likely due to the lack of affinity of the unpurified antibody (Fig. 26 A). We therefore created a lentiviral construct coding for N-terminally haemagglutinin (HA)-tagged TOX3 under the control of hPGK promoter (pER-47). The TOX3-deficient luminal breast cancer cell line MCF-7 as well as the TOX3-expressing cell lines BT-474 and MDA-MB-361, which all expressed YBX1 (Fig. 26 B), were transduced with the HA-TOX3 lentivirus
or a GFP control virus (pER-15). Expression of TOX3 and the HA-tag was confirmed by Western blot analysis (Fig. 26 C). To investigate the association of TOX3 with YBX1, I am currently immunoprecipitating HA-TOX3 using an antibody raised against the HA-tag. Co-immunoprecipitation of YBX1 will be examined by Western blot analysis for YBX1. In addition, endogenous YBX1 will be immunoprecipitated using an YBX1 antibody and co-precipitation of ectopically expressed HA-TOX3 will be examined by Western blotting for TOX3.

Tandem affinity purification of TOX3 from HEK 293T cells provides insight into the composition of protein complexes that generally associate with TOX3. However, TAP performed in HEK 293T cells fails to provide information about mammary cell-specific protein-protein interactions. For instance, HEK 293T cells do not express CITED1 (Fig. 26 B), a factor that has recently been shown to interact with TOX3 in neurons (Dittmer et al., 2010). Since CITED1 has been suggested to play a role in breast cancer and its interaction with ESR1 has been reported in several studies, I am currently investigating if the finding by Dittmer and colleagues can be reproduced in the breast cancer cell lines MCF-7, BT-474 and MDA-MB-361, which express CITED1 at varying levels (Fig. 26 B).

Figure 26.  Immunoprecipitation of TOX3.

(A) Immunoprecipitation using the TOX3 Rh57 rabbit polyclonal antibody fails to pull down ectopically expressed TOX3. (B) Western blot analysis of YBX1 and CITED1 expression in HEK 293T and breast cancer lines. (C) Western blot analysis of the hPGK-HA-TOX3 lentiviral expression construct in MCF-7 cells.
3.2.6 Affinity purification of rabbit polyclonal TOX3 antibody

The analysis of TOX3 protein expression has been complicated by the fact that no good antibody has been, and is, readily available. Since no antibody against TOX3 was commercially available when this project was started, a rabbit polyclonal peptide antibody raised against amino acids NEEDADEANR (amino acids 217-226) contained in both isoforms of human TOX3 was custom-made (Abgent) (Fig. 27).

Two rabbits were immunised with the KLH-conjugated peptide. The corresponding protein A-purified sera were termed TOX3 Rb57 and TOX3 Rb58, respectively.

Figure 27. Epitopes recognized by TOX3 antibodies.

TOX3 isoforms 1 (NP_001073899) and 2 (NP_001139660) differ in their N-terminal amino acids (in green). Custom-made rabbit polyclonal antibodies TOX3 Rb57 and TOX3 Rb58 were raised against the sequence NEEDADEANRA (underlined), which is shared by both isoforms. For comparison, the commercially available antibody ab77432 (Abcam) was raised against amino acids AGDPASLDFAQC (underlined twice), which is contained in the TOX3 isoform 1 only. For orientation, the HMG-box core sequence (red), the consensus residues within the nuclear localisation signal that are shared by all TOX proteins (pink) and the glutamine-rich region (yellow) are shown.

To characterise the TOX3 Rb57 and Rb58 antibodies, they were titrated by Western slot blot analysis of HEK 293T cells transiently transfected with a lentiviral TOX3 expression vector (pXS-39) containing the hUbiC promoter or a control plasmid (Fig. 28). Both vectors are described in the following chapter. Overexpression of hUbiC-TOX3 resulted in several bands in the Western blot. pXS-39 contained the same
lentiviral vector backbone as the hUbiC-NTAP-TOX3 vector (pXS-96) described above, it also gave rise to a TOX3-GFP fusion construct. While this explained the high molecular weight isoforms of TOX3 seen in the Western blot, it did not appear to interfere with interpretability of the antibody titration. It remained unclear why the antibody also detected several isoforms that were smaller than the predicted size of full-length TOX3. While both undiluted antibodies had been provided at an approximately equal concentration (7.5 mg/ml for Rb57 and 7.4 mg/ml for Rb58), the TOX3 Rb57 antibody appeared to give a stronger signal (Fig. 28). I therefore chose to use the TOX3 Rb57 antibody for most following experiments. Expression of endogenous TOX3 was not detectable in control-transfected HEK 293T cells.

Figure 28. Custom-made TOX3 antibody.

To test the antibody, HEK 293T cells were calcium phosphate-transfected with the lentiviral hUbiC-TOX3 expression plasmid pXS-39. hUbiC-DsRed2 control plasmid (Ctrl, pXS-41) were used as a negative control. 48 hours post transfection, whole cell lysates were prepared by direct lysis of cells in SDS-PAGE loading buffer. To titrate the antibody, proteins were separated by SDS-PAGE using a single-well preparative comb. Following the transfer, the nitrocellulose membrane was cut into strips and probed with the TOX3 antibodies (Rb57 left panel, Rb58 right panel) at different concentrations.

The custom-made rabbit polyclonal antibodies gave a clear specific signal without obvious background signal when TOX3 was overexpressed under the control of the hUbiC promoter in transfected HEK 293T cells (Fig. 28). To test the antibody under more physiological conditions, I infected MCF-7 cells, which do not express
endogenous TOX3 (Fig. 15), with lentivirus, which expresses TOX3 under the control of the human PGK promoter (pXS-36, described in the following chapter) or the negative control vector hPGK-GUS (Duss et al., 2007). Transduction of MCF-7 cell with the hPGK-TOX3 lentiviral vector resulted in TOX3 expression at levels that more closely resembled physiological expression levels. Western blot analysis of TOX3 at low levels revealed several background bands, one of which runs exactly at the same size as TOX3, as seen in the negative control lane (Fig. 29 A, E), which complicated the interpretability of results obtained with the TOX3 Rb57 antibody. Peptide blocking using the peptide that was used to immunise the rabbits for antibody production confirmed the presence of two overlapping bands, one specific and one non-specific, were overlapping. The problem with non-specific background staining became yet more evident when the detection of low levels of TOX3 required long exposure of Western blot membranes (Fig. 29 B).

To achieve cleaner Western blots that would provide more information about TOX3 expression and generally to improve antibody specificity, I took advantage of the TAP system to affinity-purify the TOX3 Rb57 antibody. To this end, NTAP-TOX3 cell lysates were prepared in TAP lysis buffer as described for tandem affinity purification above. In modification of a classic affinity column, NTAP-TOX3 was cross-linked to streptavidin-coated magnetic beads. TOX3 Rb57 antibody was allowed to bind to NTAP-TOX3 immobilised on the magnetic beads, washed several times and subsequently eluted in several fractions (Fig. 29 C, D). To remove any antibody that interacted non-specifically with the TAP tag, purified antibody was subsequently incubated with TAP control-streptavidin beads. Western blot analysis of ectopically expressed TOX3 showed that the major part of purified antibody was eluted in the first fraction (Fig. 29 D) and that affinity purification resulted in a cleaner antibody (Fig. 29 E).
Figure 29. Affinity-purification of polyclonal rabbit TOX3 Rb57 antibody.

(A) Western blot analysis shows that the antibody detects TOX3 expressed at physiological levels. (B) High background originating from the TOX3 Rb57 antibody hampers the interpretation of TOX3 expression in several different breast cancer cell lines. (C) Coomassie blue staining shows successful cross-linking of NTAP-TOX3 to streptavidin-coated beads. The arrow indicates NTAP-TOX3 protein cross-linked to streptavidin beads after washing. (D) Western blot analysis of TOX3 expression in HEK 293T cells using purified antibody fractions. Most purified antibody is eluted in the first fraction. (E) Western blot analysis MCF-7 cells transduced with NTAP-TOX3 confirms successful purification of the TOX3 antibody. Peptide blocking of the unpurified antibody demonstrates antibody specificity despite a high background.

3.2.7 Knockdown of endogenous TOX3 using shRNAmir

As discussed above, the luminal breast cancer lines ZR-75-1, BT-474, MDA-MB-361, MDA-MB-415 and UACC-812 express endogenous TOX3 protein, albeit at levels that were difficult to detect using the available antibodies (Fig. 15 B). To confirm that the band I detected using the purified TOX3 Rb57 antibody
corresponded to endogenous TOX3, I chose to knockdown TOX3 in ZR-75-1, BT-474 and MDA-MB-361 cells using two different lentiviral shRNAmir constructs directed against the sequences GACATACTGATGACTATAA (pXS-88, shTOX1 #1) and GCCTCTCTGAGTCATAGAA (pXS-89, shTOX1 #2) present in the 3’ untranslated region of the TOX3 gene. A scrambled shRNAmir vector was used as a negative control. In the first instance, I created ZR-75-1 cell lines stably expressing the three different shRNAmir constructs. Based on Western blot analysis of TOX3 expression it was not possible to quantitate the extent of knockdown as the basal TOX3 expression levels in ZR-75-1 transduced with the control shRNAmir vector were barely detectable (data not shown). Knockdown of TOX3 had no obvious effect on ZR-75-1 cell morphology (Fig. 30 B). Therefore, I tested the lentiviral shRNAmir constructs in BT-474 and MDA-MB-231 cells, which exhibited higher relative TOX3 levels (Fig. 15 B). Similar to ZR-75-1 cells, endogenous TOX3 was barely detectable by Western blot analysis in BT-474 and MDA-MB-361. However, transduction with the shRNAmir lentiviral vectors resulted in a reduction of a single band that was thought to represent endogenous TOX3, while the intensity of the other (non-specific) bands did not change (Fig. 30 A), thus confirming that the TOX3 antibody was able to detect endogenous TOX3. Quantitative PCR analysis of TOX3 mRNA expression in BT-474 cells confirmed a reduction in TOX3 expression upon transduction with the shRNAmir vectors by approximately 50% (shTOX3-1) and 40% (shTOX3-2), respectively (Fig. 30 C).

While knockdown of TOX3 using the shRNAmir constructs confirmed that the affinity-purified TOX3 Rb57 antibody was able to detect TOX3, a more efficient knockdown of TOX3 levels would be desirable in order to investigate the effect of loss of endogenous TOX3 on the tumourigenic properties of luminal breast cancer cells in our human-in-mouse tumour model, which is described in the final chapter. Due to time constraints, I was not able to follow-up the effect of TOX3 knockdown in vitro or in vivo. However, a first experiment showed that no change of expression of the TOX3 target gene PGR (described in the following chapter) could be observed in the presence of the shRNAmir constructs targeting TOX3 in BT-474 and MDA-MB-361 cells (Fig. 30 A), suggesting that either the efficiency of the knockdown was not sufficient to have a biological effect, or that knockdown of endogenous TOX3 does in fact not have an effect that is opposite to overexpression in TOX3-deficient cells.
Knockdown of endogenous TOX3 using shRNAmir vectors.

(A) Western blot analysis of TOX3 and PGR expression levels in BT-474 and MDA-MB-361 cells. Cells were stably transduced using lentiviral shRNAmir vectors targeting two different regions in the 3’ UTR of the TOX3 gene. A scrambled shRNAmir construct was used as a control. Successfully transduced cells were puromycin-selected. The closed arrow indicates the band corresponding to endogenous TOX3, while the open arrows mark non-specific bands. (B) Quantitative PCR analysis of TOX3 expression in transduced BT-474 cells. Analysis was done in triplicate, error bars represent the standard deviation. (C) Representative images of the morphology of ZR-75-1 cells transduced with the control or TOX3 shRNAmir vectors.
3.3 Identification of TOX3 target genes and signalling pathways

To gain further insight into TOX3 function and identify target genes and signalling pathways TOX3 is involved in, I chose to perform microarray analysis of mammary epithelial cells ectopically expressing TOX3 or a control gene.

3.3.1 Lentiviral expression vectors

To study TOX3 expression, I created several lentiviral TOX3 expression vectors. As mentioned above, lentiviral vectors are an efficient way to express a gene of interest in any cell type. Lentiviruses infect both dividing and quiescent cells, which makes them particularly attractive for the transduction of non-dividing stem cells, which are generally difficult to transfect. Furthermore, since they readily integrate into the host cell’s genome, lentiviral vectors allow for easy creation of cell lines stably expressing the gene of interest. As the ultimate goal of this study was to create tumour models by quantitative transformation of normal mammary epithelial cells, lentiviral vectors were the ideal choice for transgene expression.

The TOX3 open reading frame was cloned into the lentiviral backbone pSD-69 to create a lentiviral expression vector containing the human phosphoglycerate kinase (hPGK) promoter to drive TOX3 expression (Duss et al., 2007). As mentioned above, the hPGK promoter was chosen because it allows transgene expression at physiological levels. A second TOX3 expression vector was created by cloning the TOX3 ORF into the lentiviral backbone pJH-3982, which contains the human ubiquitin C promoter (UbiC) to control the expression of the gene of interest at high levels (Fig. 31). Both backbones had also been used to create the TAP vectors as described in the previous chapter. Expression of TOX3 from the two lentiviral vectors was confirmed by transient transfection of HEK 293T cells and subsequent Western blot analysis (Fig. 20 C). As expected, TOX3 protein expressed from the hPGK vector had an apparent molecular weight of just over 60 kDa, consistent with the predicted molecular weight of 63 kDa. In contrast, the hUbiC vector was subject to alternative splicing and, like the hUbiC-NTAP-TOX3 vector, produced a TOX3-GFP fusion protein, as discussed in the previous chapter. Western blot analysis and microscopic analysis of GFP localisation in MCF-7 cells confirmed that the hUbiC-
TOX3 lentiviral plasmid encoded a TOX3-GFP fusion protein (Fig. 20 C, D). As a consequence, the hUbiC-TOX3 could not be used to study TOX3 function.

Figure 31.  TOX3 lentiviral expression vectors.
Map of the lentiviral expression vectors pXS-36 (hPGK-TOX3) and pXS-39 (hUbiC-TOX3), which were cloned to create stable TOX3 cell lines. For a legend of lentiviral features, see Figure 18.

3.3.2 Microarray analysis of TOX3 target genes

To identify TOX3 target genes and gain further insight into TOX3 function, initial expression array analysis was carried out using MCF-7 cells transduced with the hPGK-TOX3 lentiviral vector or a hPGK-GUS control vector (Duss et al., 2007). To this end, MCF-7 cells were infected at a multiplicity of 10 to ensure an infection efficiency of greater than 80%.

In addition, in what had originally been designed as a separate experiment, MCF-7 cells were transduced with the two lentiviral shRNAmir vectors or the scrambled shRNAmir control lentivirus, respectively, as described in the previous chapter. When it became evident that MCF-7 cells do not in fact express endogenous TOX3 (Fig. 14, 15), the three shRNAmir samples were included in the microarray as additional negative controls. 72 hours post infection cells were lysed for extraction of total RNA. TOX3 expression was confirmed by Western blot (Fig. 32), and microarray analysis was performed using a HumanHT-12 Expression BeadChip (Illumina).
Initial filtering of the microarray data was carried out using the linear model for microarray (limma) package. Hierarchical unsupervised clustering of the most variable gene sets showed three predominating clusters of genes (Fig. 31 A). In TOX3 transduced MCF-7 cells, proliferation genes such as MYC, CCNE1, CDC45 were upregulated, while inhibitors of proliferation such as cell cycle inhibitor cyclin-dependent kinase inhibitor 1A (CDKN1A, p21) were downregulated compared to the controls (Fig. 31 A). A second cluster of genes that was downregulated in the presence of TOX3 contained luminal differentiation genes including GATA3, PGR, insulin receptor substrate 1 (IRS1), MUC1 and GREB1 (Fig. 33 A). TOX3 expression appeared to have a very weak repressive effect on ESR1 mRNA levels. The third main cluster showed a considerable induction of interferon response genes such as OAS1 in GUS transduced cells (Fig. 33 A). Induction of the interferon response as part of the innate immune response to viral infection is well established for wild-type viruses (Haller et al., 2006). In addition, induction of the interferon response can be caused by lentiviral vectors (Pebernard and Iggo, 2004). The upregulation of interferon response genes was most prominent in the GUS control, which complicated the interpretability of the microarray results and was the reason why the shRNAmir
samples were used as additional makeshift controls. To avoid experimental artefacts related to the acute cellular response to viral infection, an inducible TOX3 lentiviral vector system is currently being created, as discussed in chapter 3.5. A conditional expression vector will allow the uncoupling of viral infection and the expression of TOX3.

Nevertheless, the result that TOX3 induced proliferation while repressing luminal differentiation genes was evident despite the interferon response artefact. To determine whether the TOX3 and control gene sets showed statistically significant differences and to obtain more biologically relevant information on the effect of TOX3 expression, I carried out gene set enrichment analysis (GSEA) of the TOX3 versus the control gene sets as well as of GUS versus all other gene sets (Subramanian et al., 2005). GSEA of the GUS controls versus all other gene sets confirmed the upregulation of interferon response genes specifically in the GUS gene set (Fig. 33 C, D).

GSEA of the TOX3 arrays confirmed the induction of proliferation genes including cyclin A (CCNA2), polo-like kinase 1 (PLK1), proliferating nuclear antigen (PCNA) and MYC and the repression of the cell cycle inhibitor CDKN1A (p21) (Fig. 31 E, G, I). Moreover, GSEA showed a statistically significant downregulation of luminal differentiation genes in TOX3 arrays cells including PGR, insulin receptor substrate 1 (IRS1), GREB1, and MUC1 (Fig. 33 F, H). Somewhat surprisingly, the classic ESRI target gene TFF1 and its relative TFF3 showed no significant difference in expression (Fig. 33 I). FOXA1 expression did not change (data not shown), while TOX3 repressed the expression of the related protein FOXC1 (Fig. 33 I). TOX3 further downregulated expression of the early response gene FOS, which is in contrast to observations by Yuan and colleagues who reported that TOX3 acted as a transcriptional activator of the FOS promoter in a calcium-dependent manner (Yuan et al., 2009). Other genes that show relatively strong downregulation in the presence of TOX3 include anterior gradient homolog 2 (AGR2), which has been implicated in hormone-dependent cancers of the breast and prostate and which is expressed in the luminal cluster in ER-positive breast cancers (Brychtova et al., 2011), the aldo-keto reductase AKR1C2 gene, which plays a role in steroid hormone metabolism (Penning and Byrns, 2009), the transcription factor CITED2, which has been implicated in tamoxifen response of estrogen receptor-positive breast cancers (van Agthoven et al., 2009). TOX3 also inhibited the molecular apocrine gene activated leukocyte cell
adhesion molecule (ALCAM) (Farmer et al., 2005), and a number of histone genes (Fig. 33 B).

Figure 33. TOX3 microarray.

(A) Unsupervised hierarchical clustering using 2631 probes. MCF-7 cells were transduced with TOX3 or the control gene GUS. Illumina HumanHT-12 Expression BeadChip array results were analysed using the limma package in R, Cluster and TreeView. Black bars mark interesting gene clusters. (B) Heat map for the top 50 ranked features in the dataset. Raw Illumina data (37,249 genes) were reduced to 7065 differentially expressed genes by excluding all genes with a standard deviation of smaller than 0.1. Duplicates were removed based on HUGO gene symbols, resulting in a final set of 5988 genes. This gene set was used for GSEA. (C) GSEA enrichment plots showing enrichment in the GUS arrays for the “Interferon any response” gene set (top panel) and the “Interferon beta response” gene set (bottom panel), respectively. (D) Heat map showing the induction of interferon response genes in the GUS arrays. (E) GSEA enrichment plot showing upregulation of proliferation genes in the TOX3 arrays. (F) GSEA enrichment plot showing downregulation of genes expressed by ER-positive breast cancer. The enrichment score (ES) is the primary result of the GSEA, which reflects the degree to which a gene set is overrepresented at the top or bottom of a ranked list of genes. The nominal p value estimates the statistical significance of the enrichment score for a single gene set. The false discovery rate (FDR) is the estimated probability that a gene set with a given normalised enrichment score represents a false positive finding. (G-H) Heat map showing the induction of proliferation genes (G) and the suppression of genes upregulated in ER-positive breast cancer (H). (I) Relative change in expression of selected genes in the TOX3 versus the control arrays.
Taken together, microarray analysis of TOX3 expression suggested that TOX3 attenuates luminal differentiation and simultaneously promotes proliferation in the context of MCF-7 luminal breast cancer cells.
3.3.3 TOX3 represses a subset of ER target genes

Since the interpretation of the microarray results was hampered by the interferon response artefact seen in the control samples, I performed Western blot analysis of a selection of candidate TOX3 target genes in MCF-7 cells transduced with a TOX3 lentiviral expression vector or a control vector (Fig. 34). In keeping with the microarray results, TOX3 expression resulted in consistent downregulation of PGR, AGR2 expression. Furthermore, TOX3 expression led to a decrease in p21 protein levels, however repression of p21 was not seen every time the experiment was repeated, which may reflect differences between individual experiments with respect to the proliferative state of the transduced cells prior to harvest. In contrast to the array results, expression of IRS1 did not appear to be decreased in the presence of TOX3. ESR1 expression appeared to be reduced in the presence of TOX3 but the effect was only weak (Fig. 34). Consistent with the microarray results, TOX3 had no significant effect on expression of the classic luminal gene FOXA1 (data not shown).

![Figure 34. TOX3 target gene expression](image)

Western blot analysis of the effect of TOX3 expression on a selection of genes identified as TOX3 target genes by microarray analysis. MCF-7 cells were infected with a TOX3 expression lentivirus or a control virus. Whole cell lysates were prepared by direct lysis in protein loading buffer at least one passage post infection. The image is representative of several independent experiments.
3.3.4 TOX3 has a global effect on transcription

Since TOX3 repressed ESR1 target genes but appeared to have only a minor effect on ESR1 expression itself, I was interested to investigate if TOX3 had a direct effect on the transcriptional regulation of ESR1 target genes.

Figure 35. TOX3 acts as a global transcriptional regulator.

MCF-7 and T-47D cells were transiently co-transfected with different firefly luciferase ERE reporter vectors and a TOX3 expression vector or a DsRed2 control vector. Cells were treated with 1 µM fulvestrant (F, black bars) or 100 nM beta-estradiol (E2, grey bars) 24 hours post transfection. Cells were harvested after an addition 24 hours and analysed for luciferase activity. (A) The Renilla luciferase pRL-TK and pRL-SV40 were tested as an internal reporter control. (B) Log-normal representation of the transcriptional activation of the pGL3-basic and ERE-TK-luc reporter vectors by TOX3 versus the DsRed2 control plasmid. (C) Linear plot of the activation of two different ERE-containing reporter vectors. The data presented are the mean values ± standard error of the mean of a representative experiment performed in triplicate. Statistical significance was determined using a Student’s t-test. Significant results (p > 0.05) indicated by one asterisk are in relation to the DsRed2 control for each condition, while two asterisks indicate significant results in relation to the fulvestrant-treated equivalents.
To this end, MCF-7 cells were transiently co-transfected with the TOX3 expression vectors or a DsRed2 control vector and an estrogen receptor response element (ERE)-containing firefly luciferase reporter plasmid. The pS2-luc reporter vector contained the proximal TFF1 gene promoter in pGL3-basic, while the ERE-TK-luc vector contained the vitellogenin ERE together with the herpes simplex virus thymidine kinase (TK) minimal promoter. To control for transfection efficiency, the pRL-SV40 Renilla luciferase vector containing the SV40 enhancer and early promoter elements or alternatively, the pRL-TK Renilla luciferase vector containing the herpes simplex virus thymidine kinase (TK) promoter were used. 24 hours post transfection, cells were treated with the estrogen receptor antagonist fulvestrant (Faslodex) or beta-estradiol. In a preliminary experiment, MCF-7 cells had been maintained in phenol red-free medium supplemented with 5% charcoal-dextran treated FBS three days prior to transfection and treatment with beta-estradiol in order to exclude potential estrogenic effects of phenol-red (Shang and Brown, 2002). However, since MCF-7 cells cultured this way displayed an abnormal morphology, treatment with the estrogen receptor antagonist fulvestrant in complete medium was chosen instead of an untreated control.

The luciferase assay showed that TOX3 had an effect on the transcriptional activity of both internal control vectors (Fig. 35 A). Furthermore, activation of the TK promoter, which is supposed to provide luciferase expression at low to moderate levels, appeared to occur in a dose-dependent manner, since the hUbiC-TOX3 lentiviral construct, which produces higher expression levels of TOX3 than the hPGK-TOX3 construct, induced a 2.5-fold higher activation of the TK promoter. As the induction of the TK promoter was not of primary interest to this study, dose-dependence was not further examined. In contrast, TOX3 appeared to repress the transcriptional activity of the strong SV40 enhancer/early promoter in the pRL-SV40 construct in MCF-7 and T-47D cells, although repression in MCF-7 cells was not statistically significant. Due to the effect that TOX3 had on the internal control vectors, Renilla luciferase activity could not be used to normalise the firefly luciferase results. Since all conditions had been carried out in triplicate with generally low standard deviations, the raw firefly results may be interpretable with some confidence.

The unnormalised results suggested that in the absence of beta-estradiol induction, TOX3 had a weak activating effect on the two ERE-containing promoters compared to the DsRed2 control plasmid, however the induction was not statistically significant
Results

(Fig. 35 B, C). As expected, both ERE-containing promoter constructs were induced upon treatment with beta-estradiol, which led to an induction by approximately 4-fold for the ERE-TK-luc promoter and approximately 3-fold for the pS2/TFF1-luc promoter in the absence of TOX3. In contrast, the induction was only approximately 2-fold in the presence of TOX3 (Fig. 35 C), suggesting that the inducibility of ERE-dependent transcription by beta-estradiol was attenuated by TOX3.

In summary, it is not possible to conclude from these findings whether TOX3 acts on ESR1 target promoters. TOX3-dependent transcriptional regulation of ERE-containing promoters may depend on other transcriptional co-factors. Furthermore, it appears possible that TOX3 has a more global effect on transcription, which is difficult to assess with a luciferase assay. A more general regulatory function of TOX3 in the nucleus would be in keeping with the notion that many HMG-box proteins function as architectural chromatin elements.

3.3.5 TOX3 localises to the nucleus

Although preliminary microarray analysis suggested a function for TOX3 in the repression of luminal differentiation and the induction of proliferation, the underlying molecular mechanism of TOX3 function remained unclear. Based on sequence predictions, TOX3 contains the HMG-box DNA binding domain as well as a putative nuclear localisation signal (NLS) N-terminally adjacent to the HMG-box (Fig. 9). The presence of both domains suggests that TOX3 is a nuclear protein, which binds to DNA. To gain further cues about TOX3 function, I was thus interested in the localisation of TOX3 within the cell. Previous experiments using the hUbiC-TOX3 constructs had shown that a C-terminal GFP moiety unintentionally fused to TOX3 was able to localise to the nucleus (Fig. 20), presumably due to the presence of the NLS in TOX3. To further examine the subcellular localisation of TOX3, a C-terminal hPGK-TOX3-GFP lentiviral construct (pXS-71) was cloned. HEK 293T cells were transiently transfected with the hPGK-TOX3-GFP construct or a control construct coding for GFP (pER-15). Cells were fixed and GFP fluorescence was examined by confocal microscopy, showing that TOX3-GFP indeed localised exclusively to the nucleus, while control GFP localised to both the nucleus and the cytoplasm (Fig. 36), supporting the notion of TOX3 as a nuclear protein. Notably, GFP fluorescence levels
were markedly lower for the TOX3-GFP fusion protein than for the GFP control. One explanation for low GFP fluorescence of the TOX3-GFP fusion construct may be post-translational regulation of TOX3 protein stability. Preliminary analysis of TOX3-GFP fluorescence in transduced MCF-7 cells following treatment with the proteasome inhibitor MG132 suggested that inhibition of proteasome-dependent protein degradation did not visibly affect TOX3-GFP expression levels (data not shown). A deletion series of TOX3 domains would provide more accurate information about whether TOX3 contains domains that negatively affect its stability.

Figure 36. TOX3 localises to the nucleus.

HEK 293T cells were transiently transfected with the hPGK-TOX3-GFP fusion construct (pXS-71) or a GFP control (pER-15). 48 hours post transfection cells were treated with 1 µM etoposide (bottom panels) or vehicle alone (top panels) for 4 hours. Cells were fixed in 3.7% formaldehyde. γ-H2AX was visualised by indirect immunofluorescent staining. Nuclei were stained using DAPI and cells were analysed using a Zeiss LSM 510 Meta confocal microscope at 63 × magnification.

A recent study has shown that the TOX family member TOX4 binds to DNA adducts caused by platinating agents, a property it shares with other HMG-box family members such as HMGB1 and human upstream binding factor (UBF) (Puch et al.,
Results

2010). Therefore, I was interested to see if TOX3 was able to bind to damaged DNA. To this end, transiently transfected HEK 293T cells were treated with etoposide, which causes DNA double strand breaks (DSBs) by inhibiting topoisomerase II. One of the first events occurring following the formation of DSBs is the phosphorylation of the histone H2A variant H2AX at Ser 139 (γ-H2AX) by one or several of the kinases ATM, ATR and DNA-PK (Downs et al., 2007). Immunofluorescent staining of γ-H2AX thus marks sites of DSBs. γ-H2AX was not detected in DMSO-treated cells in the absence of TOX3, while cells stained positive for γ-H2AX following etoposide treatment (Fig 36, first and third row from top), indicating the induction of DSBs. Interestingly, some TOX3-transfected cells exhibited γ-H2AX staining in the absence of etoposide treatment, while neighbouring untransfected cells were negative for γ-H2AX (Fig. 36, second row from top). Conversely, some TOX3-transfected cells appeared to exhibit lower γ-H2AX levels than adjacent untransfected cells following etoposide treatment (Fig. 36, bottom row), suggesting a possible correlation between TOX3 expression and the presence of DNA damage. TOX3-GFP did not appear to colocalise with γ-H2AX, although co-localisation was difficult to assess, as etoposide treatment did not result in distinct γ-H2AX-containing foci.

3.4 A role for TOX3 in lineage commitment of mammary epithelial cells

As mentioned above, the prototypic member of the TOX family has a role in lineage commitment in the adaptive and innate immune systems (Aliahmad et al., 2010; Wilkinson et al., 2002). TOX expression appears to be highly regulated, and throughout lymphocyte development, the determination of CD4+ and CD8+ cell fate is controlled by changes in TOX expression (Wilkinson et al., 2002). It became clear in the course of this study that the analysis of TOX3 expression is challenging because both unsorted normal mammary epithelial cells and established breast cancer cell lines do not express abundant amounts of TOX3 protein. This suggests that endogenous TOX3, much like its relative TOX, may be subject to complex regulation. Therefore, the function of TOX3 in the regulation of mammary epithelial cell lineage commitment might best be assessed by functional assays. The basic experimental set-up was the transduction of normal mammary epithelial cells (HMEC/BPEC), the
bipotent mammary epithelial cell line MCF-10A or established breast cancer cells with the hPGK-TOX3 lentiviral expression construct or a lentiviral control construct to create cell lines that stably express TOX3. The transduced cell lines were then characterised using the mammosphere, CFC and TDLU assays described in the Introduction and below (Fig. 37).

Prior to characterising the phenotype of TOX3-expressing cells, different conditions for the culture of normal human primary mammary epithelial cells (HMEC/BPEC) were tested. As mentioned above, the choice of cell culture system for the propagation of normal and transduced mammary epithelial cells depends on the experimental goal.

![Figure 37](image.png)

**Figure 37.** *In vitro* stem cell and progenitor assays.

Schematic of the different *in vitro* cell culture assays used to characterise the phenotype of TOX3-expressing cells. The assays are described in detail in the text. MSP, mammosphere; CFC, colony-forming cell; TDLU, terminal ductal lobular unit.

### 3.4.1 Culture of human mammary epithelial cells from organoids

For all protocols involving the culture of HMEC and BPEC, tissue was obtained from reduction mammoplasties. Tissue samples were examined by a pathologist to exclude
any malignancies and pared down to obtain the ductal tissue, which was subsequently subjected to collagenase digestion to obtain organoids.

The simplest way to culture primary mammary epithelial cells is to directly place the organoids on tissue culture dishes. Four to seven days after mammary organoids were plated on Primaria dishes in WIT medium (the derived cells were therefore termed BPEC, as described above) (Ince et al., 2007), outgrowth of a monolayer of cells was observed (Fig. 38). Initial outgrowths of cells from organoids generally grew at a higher rate than cells plated from a single-cell suspension, suggesting that the microenvironment provided by the proximity of the organoid structure was beneficial to primary cell growth. This was likely due to the presence of stromal components including fibroblasts, adipocytes, endothelial cells and diverse haemapoietic cell types. In addition, the initial high growth rate of organoid-derived BPEC (compared to their dissociated counterparts) indicates that the enzymatic and mechanical dissociation conditions that cells in single suspension undergo limit their viability and proliferative potential. BPEC originating from organoids formed mixed colonies consisting of an inner core of luminal cells and a surrounding ring of myoepithelial cells when growing in a monolayer, similar to cells from single cell suspensions (described below). After one to two passages, BPEC from organoid-derived monolayer cultures resembled cells from single cell suspensions with respect to their population doubling time and morphology (Fig. 38). While organoid culture proved to be a useful way to propagate and expand human mammary epithelial cells at low passages, organoids were not suited for viral infections or quantitative assays.
Figure 38. Organoid culture of human mammary epithelial cells.

Ductal tissue from reduction mammoplasties (XS06) was mechanically dissociated and subsequently digested with collagenase to obtain organoids, which were cultured directly in WIT medium on Primaria cell culture dishes. Images are representative of the morphology of organoids and outgrowing cells at day 6 of passages 1 and 2.

3.4.2 Propagation of stem and progenitor cells in mammosphere culture

As discussed in the introduction, mammosphere culture of freshly dissociated and unsorted human mammary epithelial cells is thought to enrich for mammary stem and early progenitor cells that are thought to survive and divide in suspension culture while more differentiated epithelial cells undergo anoikis (Dontu et al., 2003). While the mammosphere culture model is arguably a less specific and controlled method for the isolation of stem and progenitor cells than cell sorting using stem and progenitor markers such as CD10, CD49f or ALDH (Bachelard-Cascales et al., 2010; Ginestier et al., 2007; Stingl et al., 2006), it is technically simpler and less detrimental to the viability of the cells.
Figure 39. Propagation of stem and progenitor cells in mammosphere culture.

Representative image of morphology of mammospheres (left, bright field; right, GFP). A single cell suspension was prepared from reduction mammoplasty tissue (XS03). Cells were left to recover overnight and subsequently infected with a GFP lentiviral vector under the control of the CMV promoter. 5×10⁴/ml cells were plated in MSPM medium on ultra-low attachment dishes. Images were taken at day five.

The protocol I used for the mammosphere (MSP) culture was essentially adapted from the original protocol developed by Dontu and colleagues, with some minor modifications (Dontu et al., 2003). Single HMEC were maintained on ultra-low attachment dishes in DMEM/F12 medium supplemented with heparin, EGF and bFGF. To achieve optimal growth conditions for the formation of spheres and to avoid aggregation of neighbouring cells, HMEC were seeded at a density of not more than 5×10⁴ cells per ml, which at passage 1 gave rise to approximately 50-100 spheres with an average diameter of 100 µm after being in culture for five days (Fig. 39). To passage the spheres, they were dissociated into single cells and re-seeded after four to seven days. Starting at day two after seeding, the formation of solid spherical structures could be observed. Dissociation of mammospheres and subsequent seeding of single cells at clonal density in adherent culture gives rise to mixed luminal and myoepithelial colonies, suggesting the presence of (at least) bi-potent progenitors in mammospheres (data not shown).

HMEC could be passaged three to four times before their capacity to form mammospheres was exhausted. With every passage, the average mammosphere size and sphere-forming frequency decreased (data not shown), suggesting that the mammosphere culture conditions do not support long-term maintenance of mammary stem cells. However, one round of mammosphere suspension culture proved to be a useful tool to select against differentiated epithelial cells and stromal contaminants and for stem and progenitor cells. Brief mammosphere culture therefore also allowed
targeted introduction of a gene of interest into progenitor and/or stem cells by lentiviral infection, as discussed below. Successful lentiviral infection of mammosphere-forming cells was confirmed by transduction with the pRRLsin.ppTs.hCMV.GFPpre lentivirus (pXS-3), encoding the GFP ORF under the control of the cytomegalovirus (CMV) promoter (Fig. 39). Since my data suggested that TOX3 was implicated in mammary epithelial cell differentiation rather than stem cell maintenance, I did not use the mammosphere protocol as a quantitative in vitro stem cell assay.

3.4.3 Propagation of luminal mammary epithelial cells in adherent culture

As discussed in the introduction, long-term maintenance of normal human mammary epithelial cells in culture still poses a problem due to the influence of the cell culture conditions on the phenotype of the cultured cells and the creation of cell culture artefacts. Limiting the time of in vitro culture prior to analysis of the cellular phenotype and genotype is therefore paramount to avoid undesired differentiation of primary and lentivirally transformed cells. However, depending on the experimental set-up, a minimum amount of time of in vitro culture is often required, for instance to allow for lentiviral infection and antibiotic selection of infected cells as well as simply to expand cells for further analysis. In an attempt to minimise aforementioned cell culture artefacts, I tested various combinations of protocols for the culture of human primary mammary epithelial cells. To this end, organoids derived from reduction mammoplasty tissue were dissociated, and single cells were subjected to one passage of mammosphere culture, followed by adherent culture in HMM+ medium (the resulting cells are referred to as HMEC). Alternatively, single cells were plated in WIT medium on Primaria cell culture plastic as described by Ince and co-workers (Ince et al., 2007), with or without an initial mammosphere step. Outgrowing cells that formed a monolayer were termed breast primary epithelial cells (BPEC). Both cell populations were maintained at 5% CO₂ and 5% O₂. It has been suggested that low oxygen levels benefits the growth of primary mammary epithelial cells, although no obvious morphological differences were observed (data not shown).
Under conventional conditions, HMEC at early passages gave rise to luminal, myoepithelial and mixed colonies (Fig. 40 A, G, H), which can be distinguished by their morphology. Luminal epithelial cells grew in colonies of tightly packed, flat cells that were generally surrounded by myoepithelial cells, which in turn were characterised by a rounder morphology and refractile edges. Myoepithelial and luminal differentiation was confirmed by immunofluorescent staining with the myoepithelial marker K14 and the luminal marker K18, respectively (Fig. 40 G, H). Since cells were plated at clonal density, I concluded that the presence of luminal, myoepithelial and mixed colonies reflected whether the cell of origin was bipotent or restricted to either mammary epithelial lineage. However, it was not possible to deduce if mixed colonies arose from multipotent mammary stem cells or bi-potent
progenitors. K14/K18 double-stained HMEC were not observed, which may be explained by the supposed low ratio of immature to differentiated cells in breast tissue, or alternatively may indicate that stem and progenitor cells do not express both keratins. Already at passage three HMEC were predominantly of myoepithelial differentiation (Fig. 40 G), suggesting that the HMM+ culture conditions do not allow the maintenance of luminal cells or strongly favour the proliferation of myoepithelial cells. HMEC typically divided for four to five passages (Fig. 40 A-C), after which they stopped dividing and exhibited a spread-out, senescent morphology (Fig. 40 B, C), suggesting that the cell culture conditions forced them to undergo replicative arrest.

In contrast, I was able to maintain BPEC in WIT medium on Primaria plates over more than ten passages without apparent induction of terminal replicative arrest (Fig. 40 D-F, I, Fig. 41 C), indicating that the WIT medium provides more appropriate conditions that favour more long-term replication and growth. Like HMEC, BPEC derived from single cell suspensions gave rise to mixed colonies (Fig. 40 D) at early passages, while BPEC derived from organoids predominantly grew in populations with myoepithelial morphology (Fig. 40 E), which was the overall phenotype at higher passages. Notably, BPEC XS08 divided with an average population doubling time of 1 doubling/24 hours at early passages and reached a plateau of 0.2 doubling/24 hours around passage nine (Fig. 41 C), which was accompanied by a somewhat senescent cell morphology (data not shown). However, at passages ten and eleven BPEC colonies emerged that appeared to resume replication at a higher rate (Fig. 41 C).

3.4.4 BPEC culture does not prevent the induction of p16

To address the observation that the primary XS08 BPEC line appeared to have undergone transient replicative arrest, I analysed the p16 status of these cells. Induction of p16 expression is thought to cause the growth arrest that is commonly seen in HMEC maintained in conventional culture conditions. A subset of HMEC in culture overcomes this arrest by silencing the p16^{INK4A} promoter by methylation (Holst et al., 2003; Romanov et al., 2001; Tlsty et al., 2004; Yaswen and Stampfer, 2001). Western blot analysis of BPEC whole cell lysates showed p16 expression throughout passages seven to eleven (Fig. 41 A). My finding was consistent with
Results

Reports that show that induction of p16 expression commonly occurs in HMEC at higher passages and coincides with positive staining for senescence-associated β-galactosidase, marking the onset of senescence (Brenner et al., 1998). Expression of p16 would indeed explain the decreased proliferation rate and senescent phenotype that was evident at passage 9. In contrast, Ince and colleagues observed p16 induction exclusively in HMEC but not in BPEC, thereby explaining why BPEC are able to proliferate indefinitely. Similarly, they reported p53 induction in HMEC but not in BPEC (Ince et al., 2007), while I saw both p53 and p21 expression in BPEC. I did not further confirm my findings using other BPEC lines.

Figure 41. Growth properties of breast primary epithelial cells.

Single primary breast epithelial cells were obtained after dissociation of fresh mammary tissue (XS08). Cells were maintained in WIT medium on Primaria plastic and serially passaged when they reached approximately 80% confluency. Whole cell lysates for Western blot analysis were prepared directly in SDS-PAGE loading buffer. (A) BPEC XS08 express p16 at passages 6-11. MDA-MB-231 breast carcinoma cells were used as a negative control. Ponceau S stain shows protein loading. (B) BPEC XS08 express p53 and p21 at passages 6 and 7. (C) Population doubling times at serial passages.

These somewhat contradictory findings underline once more the caveats of long-term culture of primary mammary cells and emphasise that it is crucial to limit the time of in vitro culture until a reliable culture system has been developed. However, variations in the growth rate and morphology of BPEC may also be explained by differences between mammoplasty samples such as the age of the patient or the time required to process the sample. Another possible explanation may be that the cell culture conditions varied between laboratories due to the complexity of the WIT medium.
Nonetheless, comparison of BPEC and HMEC growth properties demonstrated that BPEC could be maintained in culture longer without undergoing senescence. In addition, using gene expression profiling of HMEC and BPEC, Ince and colleagues showed in their original study that culture in WIT medium favoured luminal differentiation of cells, while HMEC overexpressed more than twice as many myoepithelial genes as BPEC (Ince et al., 2007). Since my working hypothesis was that TOX3 was a regulator of luminal differentiation, I chose the BPEC/WIT protocol for further experiments with the intention to create a physiological environment for TOX3 function.

### 3.4.5 TOX3 expression decreases the myoepithelial subpopulation

To examine whether TOX3 had an effect on mammary lineage commitment, TOX3 was expressed in non-transformed MCF-10A mammary epithelial cells. MCF-10A cells were chosen as a model system for initial experiments as they exhibit bipotent progenitor-like properties and recapitulate epithelial morphogenesis in three-dimensional culture by forming acinar structures (Debnath and Brugge, 2005; Debnath et al., 2003; Herr et al., 2011), while their long-term culture is less challenging than that of primary mammary epithelial cells. The differentiation state of transduced cells was analysed using the colony-forming cell (CFC) assay as described by Bachelard-Cascales and co-workers (Fig. 37) (Bachelard-Cascales et al., 2010). MCF-10A cells plated at clonal density on a layer of irradiated mouse embryonic fibroblast (MEF) feeder cells gave rise to myoepithelial, luminal and mixed colonies, thus behaving like bipotent mammary epithelial progenitor cells (Fig. 42 A). To quantify the effect of TOX3 expression on MCF-10A differentiation, 2000 cells were transduced with the hPGK-TOX3 lentiviral vector or the hPGK-DsRed2 control vector. TOX3 expression was confirmed by Western blot analysis (Fig. 42 B), and colony formation was examined after seven days (Fig. 42 C). Cells expressing DsRed2 showed an abnormal morphology, suggesting that DsRed2 expression had a toxic effect (data not shown). Therefore, a no virus control was included in the analysis. Both DsRed2 and non-transduced cells gave rise to approximately five-fold more myoepithelial colonies than cells expressing TOX3, suggesting that TOX3 inhibited differentiation of MCF-10A cells along the myoepithelial lineage. Cells
expressing TOX3 gave rise to approximately six times more luminal colonies than the DsRed2 control but the number of luminal colonies did not change significantly compared to non-transduced cells (Fig. 42 C). FACS analysis of the phenotype of TOX3 expressing cells was somewhat inconclusive. While TOX3 expression resulted in a decrease of the CD24+ subcompartment, it led to an increase in the percentage of EpCAM+ cells. Since both are markers of luminal differentiation, the significance of these observations was unclear. Interestingly, TOX3 expression resulted in a reduction of the CD10+ compartment, which was consistent with a potential inhibition of cells with a myoepithelial phenotype (Fig. 42 D), as seen in the CFC assay. Furthermore, TOX3 expression resulted in an increase of the CD10/EpCAM double-positive subpopulation, while it decreased the CD10/CD49d double-positive subcompartment, which are thought to be enriched for luminal and myoepithelial progenitors, respectively (Fig. 42 D).

It is tempting to speculate that TOX3 expression at an early stage of mammary epithelial development results in a decrease of the early progenitor population by forcing differentiation of MCF-10A cells along the luminal lineage, which in turn would result in an increased percentage of committed luminal progenitors and a simultaneous decrease of the committed myoepithelial progenitor and mature myoepithelial subpopulations. However, the CFC results and the phenotypic characterisation have to be interpreted with caution since DsRed2 expression was toxic for the control cells and the nontransduced cells are not a fully adequate control. The assays therefore need to be repeated with an appropriate lentiviral control vector.
Results

Figure 42. TOX3 expression inhibits myoepithelial differentiation.

(A) Representative images of colony-forming cell (CFC) assay read-out. MCF-10A cells were plated on a feeder layer of irradiated mouse embryonic fibroblasts (MEF). After seven days, cells were fixed in ice-cold methanol and stained with May-Grünwald colorant. (B-C) To assess the effect TOX3 on mammary epithelial cell differentiation, MCF-10A cells were infected with hPGK-TOX3 lentiviral vector or hPGK-DsRed2 lentiviral control vector at a multiplicity of infection of 20. 2000 cells were seeded on a MEF feeder layer. Colony formation was analysed at day 7 following infection. Non-transduced (no virus, NV) cells were included as a second control because DsRed2 overexpression appeared to be toxic to the cells. (B) Western blot analysis of TOX3 expression at day 7. (C) CFC assay. (D) FACS analysis of MCF-10A cells at day 7 after transduction with TOX3 lentivirus.
3.4.6 Expression of TOX3 does not change ESR1 expression in BPEC

Due to time constraints I was not able to confirm the results of the CFC assay using freshly dissociated BPEC stably expressing TOX3. Preliminary analysis showed no obvious morphological difference between BPEC transduced with TOX3 or the GUS lentiviral control vector (Fig. 43). In contrast, expression of the estrogen receptor α (ESR1) quickly led to the induction of terminal differentiation in BPEC, resulting in a senescent, flattened phenotype (Fig. 43 C). The different phenotypes of cells expressing TOX3 or ESR1 were consistent with the MCF-7 microarray data discussed above. Western blot analysis confirmed that TOX3 expression in BPEC did not induce ESR1 expression, which was undetectable in control and TOX3-expressing cells (Fig. 43 A).

Figure 43. TOX3 expression in breast primary epithelial cells (BPEC).

BPEC XS11 were transduced with lentiviruses coding for GUS, TOX3, or ESR1, respectively, at passage 3. (A) Western blot analysis of transduced BPEC. TOX3 does not alter ESR1 expression levels. (B) Representative phase contrast images of the morphology of BPEC expressing GUS, ESR1 or TOX3, respectively, at passage 4 (magnification 20 ×).
3.4.7 Three-dimensional culture of TOX3-expressing cells

As discussed above, in vivo reconstitution assays are the ultimate stem cell assay. An in vitro alternative is what has been termed the terminal ductal lobular unit (TDLU) assay, which is used to evaluate the capacity of cells to give rise to three-dimensional TDLU-like structures when embedded in a matrigel matrix (Bachelard-Cascales et al., 2010). To investigate if TOX3 alters the property of cells to produce three-dimensional structures, BPEC and MCF-10A cells transduced with TOX3 (pXS-36) or a lentiviral control vector coding for DsRed2 (pSD-136) or GFP (pER-15) were seeded at a single injection site in matrigel mixed with MEF feeder cells and maintained in WIT medium. In addition, the luminal breast cancer cell line MCF-7 and the basal breast cancer cell line MDA-MB-231 were examined. For the successful outgrowth of tumour cells feeder cells were not required. While the TDLU assay had originally been developed to assess stem cell properties of primary mammary epithelial cells (Villadsen et al., 2007), my rationale was that TOX3 might alter the differentiation of established tumour cells that do not express endogenous TOX3.

Three-dimensional growth was examined after 6 and 14 days. As observed in the CFC assay, expression of the fluorescent control genes DsRed2 (multiplicity of infection, MOI = 20) and GFP (MOI = 5) appeared to be toxic to the cells and impaired cell growth, as assessed by comparison to the no-virus control (NV). This effect did not appear to be due to the lentiviral infection per se but rather the efficient expression of the fluorescent marker genes at relatively high levels, even at low MOIs, or an underestimate of the true viral titre. TOX3-expressing cells were therefore compared to both negative controls. As expected, bipotent MCF-10A cells gave rise to three-dimensional outgrowths that resembled duct-like and bulbous structures (Fig. 44). Microscopical analysis did not show significant morphological differences between outgrowths derived from TOX3- and control-infected cells but without histological analysis the formation of TDLU-like structures is difficult to interpret. To examine their differentiation state, three-dimensional outgrowths are currently being examined by immunohistochemical staining for keratins 14 and 18. While the TDLU assay confirmed the bipotent properties of the MCF-10A cell line, it gave no valuable information on the effect of TOX3 expression in the MCF-7 and MDA-MB-231 tumour cell lines. MCF-7 cells in particular were not able to grow out in duct- and TDLU-like structures but rather formed a solid lump at the site of injection, and no
obvious morphological difference could be observed between TOX3-expressing and control cells (data not shown). MDA-MB-231 cells appeared to be able to grow out in the matrigel matrix to a greater extent. Control-infected cells formed thin tubular structures but no TDLU-like outgrowths (data not shown), while TOX3-transduced cells remained in a solid lump at the injection site, suggesting that TOX3 repressed growth of the basal tumour cell line but again, more sound conclusions can only be drawn upon phenotypic characterisation of the differentiation state of the outgrowths. Due to technical problems, characterisation of TOX3-expressing primary BPEC in the TDLU assay could not be concluded but is subject of current investigations.

![MCF-10A cells](image)

Figure 44. TDLU assay.

MCF-10A cells were trypsinised and infected with hPGK-GFP (pER-15) or hPGK-TOX3 lentiviral vectors at a MOI of 5 and 20, respectively. In addition, a no-virus control (NV) was included. Cells were subsequently injected into hardening matrigel mixed with MEF on a Lab-Tek chamber slide. Hardened matrigel was covered with medium. Outgrowths were examined at days 6 (top row) and 14 (bottom row).

### 3.5 A TOX3 tumour model

To test my working hypothesis that TOX3 is a transforming oncogene, I set out to create a luminal xenograft tumour model based on quantitative transformation of normal human mammary epithelial cells, as described in the Introduction. As an initial experiment, I chose to examine the effect of stable TOX3 expression in the luminal tumour cell line MCF-7. The rationale for using an established cell line was
twofold. First, microarray analysis of short-term TOX3 expression in MCF-7 cells had identified several TOX3 target genes. Secondly, the effect of TOX3 expression on the morphology and differentiation of normal human mammary epithelial cells and non-tumourigenic MCF-10A cell growth appeared rather subtle, suggesting that expression of TOX3 alone was not able to transform non-tumourigenic cells. MCF-7 cells were chosen as a first model system as they are tumourigenic cells of luminal differentiation, thereby providing an appropriate environment for TOX3 function.

3.5.1 TOX3 expression confers short-term growth advantage but does not alter long-term proliferation of MCF-7 cells

To analyse stable TOX3 expression, MCF-7 cells were infected with hPGK-TOX3 lentivirus (pXS-36) or a negative control vector expressing DsRed2 (pSD-136) or GFP (pER-15) at a multiplicity of infection that ensured a transduction efficiency of greater than 80%. Transduced cells were selected by puromycin treatment. In addition to MCF-7 cells, TOX3 was expressed in the basal breast cancer cell line MDA-MB-231 to investigate whether TOX3 would alter the growth properties of basal tumour cells. TOX3 expression was examined by Western blot at day 7, 11, 15, and 22 post infection.

While TOX3 expression did not have any obvious effect on the morphology of either MCF-7 or MDA-MB-231 cells (data not shown), Western blot analysis confirmed TOX3 expression but also indicated that expression was silenced in MCF-7 cells over time (Fig. 45 A). Several repeats of hPGK-lentivirus-based TOX3 expression in MCF-7 cells as well as HMEC AJ4 and T-47D cells confirmed that after approximately three weeks (or 6 to 7 passages), relative TOX3 expression was markedly decreased (data not shown). To date, I have not been able to find a conclusive explanation for the silencing of TOX3 expression. The microarray results suggested that TOX3 weakly repressed PGK expression (Table 14, Appendix), which would create a negative feedback loop. The silencing of TOX3 expression using our lentiviral model system posed a major problem for the creation of stable TOX3 cell lines and as a consequence also for the creation of a TOX3 tumour model.
Figure 45. Stable expression of TOX3 in MCF-7 and MDA-MB-231 cell.

MCF-7 cells were transduced with hPGK-TOX3 (pXS-36) or hPGK-DsRed2 control vector and selected using puromycin. (A) Expression of TOX3 and its target gene PGR at day 7, 11, 15, and 22 post infection was examined by Western blot analysis. (B) TOX3 had no significant immediate effect on the proliferation rate of MCF-7 cells. Cells were counted at day 1-7 following passage 1 post infection. The experiment was repeated three times. A representative growth curve is shown. (C) Effect of long-term TOX3 expression on the proliferation rate of MCF-7 cells. Transduced cells were maintained in culture for 21 days and passaged every 3 days. The same number of cells was seeded and cells were counted at every passage. The experiment was repeated three times. A representative growth curve is shown. Error bars represent the standard deviation.

To circumvent the use of the constitutive hPGK-TOX3 expression vector, we are currently creating a drug-controllable lentiviral vector for inducible TOX3 expression using the KRAB repressor/tet-off system (Szulc et al., 2006). An inducible system will allow tight control of TOX3 expression. In addition, it will be possible to uncouple the time of lentiviral infection and the time of TOX3 expression, thus also temporally separating potential side-effects of lentiviral infection such as the
Results

induction of the interferon response from TOX3-specific effects. Without an inducible TOX3 expression vector at hand, analysis of TOX3 expression had to be limited to approximately three weeks. During this period of time, TOX3 function could be confirmed by downregulation of the TOX3 target gene progesterone receptor (Fig. 45 A).

To examine whether TOX3 expression affected the proliferation rate of MCF-7 cells, I first followed cell growth over the course of seven days following the first passage post infection. TOX3 appeared to have no significant effect on the proliferation rate (Fig. 45 B). To examine whether TOX3 expression would have a more obvious effect over a longer period of time, MCF-7 cell proliferation was followed for approximately three weeks following the first passage after lentiviral infection. The proliferation rate was consistently increased at passage 2 post infection (equalling approximately one week) in the presence of TOX3 but the effect was no longer detectable at the following passage (1.5 weeks), despite continued TOX3 expression (Fig. 45 A), suggesting that the onset of TOX3 expression induced proliferation but sustained TOX3 expression did not maintain this effect. Preliminary cell cycle analysis using flow cytometry showed a corresponding transient small increase in the percentage of cells in S phase at passage 2 following infection in the presence of TOX3 (data not shown). However, albeit being detectable in several separate experiments, the increase was not statistically significant.

3.5.2 TOX3 increases the tumour take in the MCF-7 luminal tumour model

To test the tumourigenic potential of TOX3, MCF-7 cells expressing TOX3 or the control gene were xenografted in 6-week old female NSG mice by intraductal injection (Behbod et al., 2009). Tumour formation was followed by in vivo fluorescence imaging for 21 days, and mice were sacrificed and examined for tumour incidence at days 7 and 21 post injection. All control (8/8 mice) and TOX3 mice (11/11) developed tumours, demonstrating that both control and TOX3-expressing MCF-7 cells were able to engraft in the murine ducts, albeit with different efficiency. At day 7, the tumour load in TOX3 mice was significantly higher than in control mice (Fig. 46 A, B). However, at day 21, no difference in tumour load could be detected between TOX3 and control mice, suggesting that TOX3 increased the initial tumour
take in the MCF-7 luminal tumour model. It was not clear whether this was due to increased proliferation of the TOX3-expressing cells or whether TOX3 promoted engraftment of the cells in the ducts. Histological analysis showed that tumours were initially confined to the ducts but rapidly progressed to invade the fat pad (Fig. 46 C).

TOX3 increases tumour take in the MCF-7 luminal breast cancer model. MCF-7 cells were infected with a GFP lentiviral vector not conferring puromycin resistance. Subsequently cells were superinfected with hPGK-TOX3 or DsRed2 lentivirus and subjected to puromycin selection. Cells were intraductally injected into the mammary gland of 6-8 week-old NSG mice and sacrificed at day 7. (A-B) Representative images of GFP fluorescence of MCF-7 cells in TOX3 mice (A) and control mice (B) at day 7. Tumour load in TOX3 mice was significantly higher than in control mice. (C) H&E staining of TOX3-expressing MCF-7 cells showing the presence of carcinoma invading the surrounding adipose tissue. (D) Immunohistochemical (IHC) staining for GFP confirming that the observed tumours were formed by injected MCF-7 cells. (E) IHC staining for estrogen receptor α.
3.5.3 Quantitative transformation of normal mammary epithelial cells

In view of the problems related to the creation of breast cancer cell lines stably expressing TOX3, the technically more challenging transformation of normal mammary epithelial cells did not seem advisable without the creation of a conditional TOX3 lentiviral expression construct. As mentioned above, we are currently creating a drug-controllable lentiviral vector for inducible TOX3 expression using the KRAB repressor/tet-off system (Szulc et al., 2006), which will allow tightly controlled TOX3 expression.

While a reliable TOX3 lentiviral expression construct was unavailable, I carried out a small pilot experiment to establish a transformation protocol for normal human mammary epithelial cells. The protocol I decided to use was adapted from the luminal orthotopic xenograft tumour model that had previously been developed in our laboratory (Duss et al., 2007). In contrast to the original model, which used HMM+ medium for the culture of transformed HMEC on conventional plastic dishes, I chose to maintain cells in WIT medium on Primaria plastic. As discussed above, Ince and colleagues had shown that the latter culture method promoted luminal differentiation of primary mammary epithelial cells and prevented squamous differentiation of BPEC-derived tumours in immunodeficient mice, which is a problem commonly observed with HMEC-derived xenografts (Duss et al., 2007; Ince et al., 2007).

For the transformation protocol (Fig. 47), mammary epithelial cells (BPEC) and human mammary fibroblasts (HMF) were obtained from reduction mammoplasties of healthy pre-menopausal women. Dissociated HMF were plated directly in adherent culture. Following dissociation, single BPEC were allowed to recover overnight in suspension before they were infected with lentiviruses encoding estrogen receptor α (hPGK-ESR1, pSD-82) and BMI1 (hPGK-BMI1, pSD-84). Negative control cells were infected with an hPGK-GFP lentiviral vector (pRRlsin.ppts.hPGKGFPpre, pXS-17). Infected cells were cultured as mammospheres for one passage to enrich for progenitor cells. Subsequently, cells were plated in adherent culture, and successfully transduced cells were selected using puromycin.
Results

Figure 47. Protocol for the creation of a luminal tumour model.
Schematic of the protocol used to create a luminal tumour model using the quantitative transformation BPEC and subsequent xenograft to test tumourigenic potential of transformed cells.

In a second round, BMI1/ESR1-positive cells were infected with lentiviruses coding for MYC (hPGK-MYC, pSD-94) and TERT (hPGK-TERT, pSD-83). In addition, cells were infected with the hPGK-DsRed2 lentivirus (pSD-136) for subsequent tracking in the mouse. DsRed2 transduction was confirmed by fluorescence microscopy (data not shown), while expression of the remaining transgenes was examined by Western blot (Fig. 48 A). While the expression of ESR1 and BMI1 was confirmed, TERT and MYC expression were only detected at very low levels. As all lentiviruses shared the pSD-69 lentivirus backbone that conferred resistance to puromycin, another selection step was not possible to ensure higher levels of TERT and MYC. To test their tumourigenic properties, 75,000 BPEC XS08.4 cells (passage 8) expressing the four transgenes plus the DsRed2 marker were unilaterally co-injected with 25,000 HMF mixed with 12.5 vol% matrigel into the fourth inguinal mammary gland of 30-week old female NSG mice one week post insertion of a 0.5-mg slow-release beta-estradiol pellet. GFP-expressing BPEC (passage 6) were used as
a negative control, and HMEC AJ4 cells (passage 16) were used as a positive control. Mice were sacrificed after 15 days and analysed for tumour and metastasis incidence. The results are summarised in Table 10.

<table>
<thead>
<tr>
<th>BPEC/HMEC line</th>
<th>Colour marker</th>
<th>Tumours</th>
<th>Metastases</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMEC AJ4 (P16)</td>
<td>CFP</td>
<td>1/1</td>
<td>0/1</td>
</tr>
<tr>
<td>BPEC XS08.4 (M1P7)</td>
<td>DsRed2</td>
<td>2/3</td>
<td>0/3</td>
</tr>
<tr>
<td>BPEC XS08.ctrl (M1P5)</td>
<td>GFP</td>
<td>0/1</td>
<td>0/1</td>
</tr>
</tbody>
</table>

HMEC AJ4 and BPEC XS08.4 expressed ESR1, BMI1, TERT, and MYC; the control line BPEC XS08.ctrl expressed GFP. The passage number of injected cells is indicated in parentheses; P, adherent passage; M, mammosphere passage.

As expected, the negative control cell line BPEC XS08.ctrl did not induce tumour formation in mice. Consistent with previous results from our group, AJ4 cells were able to form tumours in mice (Duss et al., 2007). However, they did not form metastases as had been observed previously, which was likely to be due to the relatively high passage number (Stephan Duss, personal communication). BPEC XS08.4 cells were able to form tumours in two out of three cases (Table 10, Fig. 48). Inconsistent with the reports from the Weinberg laboratory (Ince et al., 2007; McAllister et al., 2008), tumours were predominantly of squamous differentiation (Fig. 47 D, E). BPEC formed squamous tumours with an efficiency similar to the HMEC AJ4 line (data not shown). Consistent with the findings of Weinberg and colleagues, I observed a strong stromal reaction surrounding the BPEC-derived tumour tissue (Fig. 48 D, E), suggesting that BPEC XS08.4 cells represent instigating tumour cells capable of recruiting activated murine stromal cells by secreting osteopontin (McAllister et al., 2008).
Results

Figure 48. Quantitatively transformed BPEC form tumours in NSG mice.
(A) Western blot analysis of transgene expression. The HMEC tumour cell line AJ4 was used as a positive control (Duss et al., 2007). (B) Phase contrast image of tumour at day 15. (C) DsRed2 fluorescent image of the same tumour area at day 15. (D) H&E staining of tumour tissue demonstrates squamous differentiation. Cross-section through tubular tumour structure. (E) Immunohistochemical staining of tumour cross-section for cytokeratin 14 (brown) confirms the presence of squamous differentiation. The stained tumour tissue is surrounded by an area of extensive stromal reaction.

In conclusion, in my hands BPEC-derived tumour cells exhibit some of the characteristics that Weinberg and colleagues described but differ in that they are not capable of forming adenocarcinomas. In addition, untransformed BPEC display different growth and differentiation properties than the cells in the original study, as described in the previous chapter. It has to be noted that a larger number of mice was not available at the time of the experiment, which impairs the statistical relevance of the data I obtained from this preliminary experiment. However, BPEC could be oncogenically transformed using a combination of TERT, MYC, BMI1 and ESR1. I was thus able to confirm the tumourigenic properties of quantitatively transformed mammary epithelial cells in the xenograft model. All tools are thus set up to test the tumourigenic potential of TOX3 once the inducible BPEC-derived cell lines have been created.
4 Discussion

Estrogen receptor α (ERα, ESR1) expression is a major determinant of the breast cancer phenotype and biology. As a result, the mechanisms underlying ERα-positive breast cancer have been the subject of extensive research. However, to date many questions have remained unanswered including the paradox that estrogen acts as a potent mitogen in ERα-positive breast cancer but induces differentiation in normal mammary epithelial cells. The identification of new luminal breast cancer genes and the establishment of model systems that accurately mimic the human disease and allow the characterisation of these genes are therefore of great interest.

In the present study, I have investigated the function of the previously uncharacterised putative breast cancer predisposition gene TOX3 in the normal breast and ERα-positive breast cancer. The broader context of this work was the characterisation of new candidate breast cancer genes in the ERα-positive human-in-mouse tumour model that had previously been established in our laboratory (Duss et al., 2007). The latter model was developed by transformation of normal human mammary epithelial cells with the genes TERT, BMI1, MYC and ESR1. The resulting cells gave rise to ERα-positive tumours with high efficiency and were estrogen-dependent for growth. SNP chip analysis showed that the human tumour cells were genetically normal after growth in the mouse, suggesting that transformation was caused by expression of the four transgenes alone (therefore termed “quantitative” transformation). In addition, the cells formed metastases, thus displaying several properties of human ERα-positive breast cancer. However, histological examination of the tumour tissue showed large regions of squamous differentiation, which is rare in human ERα-positive tumours. While the formation of squamous carcinoma has been linked to inappropriate cell culture conditions (Ince et al., 2007), another explanation may be that the combination of transgenes used to create the model did not adequately reflect the genetic changes causing the human disease, thus inducing an aberrant transcriptional programme, which results in differentiation into keratinised epithelial cells. It appears likely that ESR1 along with BMI1 and MYC caused the formation of tumours but the correct cellular differentiation and thus development of adenocarcinoma may require the
expression of additional co-regulators of ESR1. The rationale of the present work was therefore to characterise such potential co-factors of ESR1.

TOX3 was chosen as a candidate gene because it clusters with classic luminal genes such as GATA3, FOXA1, and XBP1 in expression array data of primary breast tumours previously described by our group (Farmer et al., 2005). The fact that augmented TOX3 expression correlates with expression of the luminal differentiation genes in ERα-positive suggests that TOX3 may be subject to a regulatory mechanism that is similar to that of FOXA1 and GATA3 and that TOX3 may have a related function. In addition, recent genome-wide association studies had linked single nucleotide polymorphisms in the regulatory region of the TOX3 gene to an increased susceptibility to ERα-positive breast cancer (Easton et al., 2007; Stacey et al., 2007).

As TOX3 gene function was entirely uncharacterised at the time of the start of this project, the present study was designed as a broad approach to investigate the biological role of TOX3.

4.1 TOX3 is a luminal gene

In the normal mammary epithelium overall expression of TOX3 protein was comparatively low, as it was not detectable by Western blot analysis of unsorted BPEC (Fig. 16 A). Due to the unavailability of sufficient amounts of fresh reduction mammoplasty tissue I was not able to investigate TOX3 expression in FACS-sorted mammary epithelial cells by quantitative PCR analysis myself. However, expression data were available from two recent microarray studies from the Smalley and Visvader groups (Kendrick et al., 2008; Lim et al., 2010). Based on these data, TOX3 is expressed almost exclusively in the mature luminal epithelial compartment but not mammary stem cells or luminal progenitors in the human breast (Fig. 16 C) (Lim et al., 2010), which may suggest that TOX3 is predominantly involved later in luminal differentiation but that it does not function as a regulator of early luminal cell fate determination. The mouse data from the Smalley group further showed relatively high expression of TOX3 in ERα-positive luminal cells versus low expression in ERα-negative luminal cells (Fig. 16 C) (Kendrick et al., 2008). The evolutionarily conserved expression pattern of TOX3 (Table 8) suggests a function late in the
differentiation of luminal epithelial cells. It remains to be determined how exactly TOX3 regulates luminal cell fate. Possible scenarios may involve active promotion of luminal differentiation (Fig. 49), which would be consistent with the observation that TOX3 expression in MCF-10A cells increased the EpCAM+ subpopulation (Fig. 42 D). Alternatively, TOX3 may favour differentiation along the luminal cell lineage by inhibiting myoepithelial cell differentiation (Fig. 49). This would be in keeping with the observation that TOX3 expression in MCF-10A cells resulted in a decrease in myoepithelial colony formation in the CFC assay (Fig. 42 B) as well as in a decrease of the CD10+ and CD49d (ITGA4)/CD10 double-positive myoepithelial subpopulations (Fig. 42 D). It is tempting to speculate that TOX3 expression in normal epithelial cells maintains luminal cell differentiation by inhibiting the commitment to a myoepithelial cell fate. This would be in analogy to the related family member TOX, which regulates the development of both CD4+ and CD8+ cells in the thymus in a lineage-specific manner (Aliahmad and Kaye, 2008; Aliahmad et al., 2004; Wilkinson et al., 2002). Such a dual role for TOX3 in mammary epithelial cell development would likely involve the interaction with other transcriptional coregulators such as YBX1 (Table 9) and CITED1, which is discussed in more detail below.

Taken together, evidence from the transcriptional profiling studies and the CFC assay suggest a role for TOX3 in the biology of luminal epithelial cells but further functional assays will be required to identify its exact function. For instance, it would be of interest to express TOX3 in the different mammary epithelial subpopulations to examine its effect on lineage choice and differentiation.

4.2 A transcriptional network that regulates luminal epithelial development and differentiation

4.2.1 TOX3 is a nuclear protein

I show here for the first time that TOX3 is a nuclear protein. An ectopically expressed TOX3-GFP fusion protein localised exclusively to the nucleus of MCF-7 and HEK 293T cells. TOX3-GFP exhibited an overall diffuse nuclear localisation that was interspersed with regions that appeared to be in more condensed, ring-like or patchy
conformations. TOX3 was further excluded from nucleoli (Fig. 20, 36). The subnuclear localisation pattern of the TOX3-GFP fusion protein bore resemblance to those of canonical HMGB proteins, which in turn have been described to recapitulate the localisation pattern of histones and other chromatin-bound proteins as well as DNA itself (Catez and Hock, 2010). Nuclear localisation of transiently and stably expressed TOX3 was confirmed by indirect immunofluorescent staining but owing to the impurity of the custom-made TOX3 Rb57 antibody, high background levels complicated the interpretation (data not shown). Nuclear localisation of TOX3 is in keeping with the predicted nuclear localisation signal adjacent to the HMG-box on its N-terminal side. Like TOX3, the related proteins TOX and TOX2 have been shown to localise to the nucleus (Aliahmad and Kaye, 2006; Kajitani et al., 2004).

Immunofluorescent analysis of TOX3-GFP and γ-H2AX co-localisation was inconclusive due to the relatively diffuse subnuclear localisation of both proteins in etoposide-treated cells (Fig. 36). γ-H2AX typically localises to distinct foci at sites of DSBs caused by ionising radiation (Nakamura et al., 2010). A possible explanation for the diffuse localisation of γ-H2AX that I observed may be that cells were treated with etoposide for four hours, while phosphorylation of H2AX occurs rapidly after the induction of DNA damage. To investigate whether TOX3 localises to γ-H2AX foci at sites of DSBs, it would perhaps be preferable to cause DNA damage using ionising radiation and to analyse subnuclear localisation of both proteins shortly after the induction of DNA damage. Surprisingly, TOX3-transfected cells stained positive for γ-H2AX, which may suggest that TOX3 plays a role in the recognition or repair of damaged DNA like the related protein TOX4 (Puch et al., 2010). Phosphorylation of H2AX in TOX3-expressing cells requires further analysis and could be verified by Western blot analysis. Expression of TOX3-GFP under the control of a stronger promoter such as the CMV promoter would hopefully produce microscopy data of higher quality that would facilitate the analysis of TOX3 localisation.

In contrast to the localisation of TOX3-GFP, IHC staining of TOX3 in luminal tumours was both cytoplasmic and nuclear (Fig. 12). The discrepancy between IHC data and localisation of TOX3-GFP may be due to different posttranslational modification of endogenous TOX3 and ectopically expressed TOX3, which may result in its localisation to the nucleus in the normal breast (or cells that do not express endogenous TOX3) but may lead to its sequestration in the cytoplasm in tumour cells. Nuclear export or translocation to the nucleus is a common mechanism
to regulate the activity of transcription factors. On the other hand, cytoplasmic staining may be an artefact and simply due to insufficient purity of the TOX3 Rb57 antibody that was used. To characterise subcellular localisation of TOX3, it would be of interest to create a deletion series of TOX3-GFP fusion constructs.

As mentioned above, various functions have been ascribed to HMG box-containing proteins that are independent of their role in classic transcriptional regulation. With respect to a potential involvement in DNA damage recognition or repair, it may be of interest that the microarray experiment showed that a number of histone genes were among the most downregulated genes in TOX3-expressing MCF-7 cells (Fig. 33 B). Histone gene expression is known to be regulated in a cell cycle-dependent manner. Therefore, the observed downregulation of histone genes may also be related to the changes in replication gene expression (Fig. 33).

### 4.2.2 Transcriptional regulators of luminal differentiation

As mentioned above, evidence for a role for TOX3 in ERα-positive breast cancer comes from the finding that TOX3 is expressed in luminal tumours but not in ERα-negative tumours and importantly, because it clusters with the genes that define the luminal breast cancer subtype (Fig. 11). Based on sequence predictions, the HMG-box motif in TOX3 is thought to bind DNA without sequence specificity, similar to a multitude of HMG-box proteins that organise chromatin structure as architectural proteins (O'Flaherty and Kaye, 2003). It has to be noted that the present study does not provide direct proof that TOX3 acts as a transcriptional regulator in mammary cells but evidence that TOX3 is a transcription factor comes from studies showing that TOX3 cooperates with CREB and CITED1 to activate transcription in neurons (Dittmer et al., 2010; Yuan et al., 2009). It is thus plausible that, like other luminal transcription factors, TOX3 plays a role both in the normal development and differentiation and in the oncogenic transformation of luminal epithelial cells. Beside the Ets transcription factor ELF5, which regulates luminal alveolar cell fate (Oakes et al., 2008a), FOXA1 and GATA3 are probably the best-characterised luminal transcription factors (Asselin-Labat et al., 2007; Bernardo et al., 2010; Kouros-Mehr et al., 2008; Kouros-Mehr et al., 2006). In keeping with their role in luminal cell differentiation, expression of FOXA1 and GATA3 in breast cancer correlates with a
Discussion

good prognosis (Badve et al., 2007; Habashy et al., 2008; Mehra et al., 2005; Sorlie et al., 2003). Although the correlation between ESR1, FOXA1 and GATA3 expression in luminal breast cancer is well established, surprisingly little is known about the interplay between these factors. A recent study showed that FOXA1, aside from its role as a co-regulator of ESR1-dependent transcription, also regulates expression of ESR1 but not of GATA3 (Bernardo et al., 2010). FOXA1 in turn has been shown to be a downstream target of GATA3 (Kouros-Mehr et al., 2006). In breast cancer cells, FOXA1 appears to mediate, at least in part, the pro-proliferative properties of ESR1 as it is required for transcriptional activation of cyclin D1 (CCND1) by ESR1 (Eeckhoute et al., 2006). Similarly, GATA3 functions in a positive cross-regulatory loop with ESR1 in breast cancer cells (Eeckhoute et al., 2007).

It is conceivable that TOX3, too, functions in the transcriptional network that controls luminal differentiation. Western blot and microarray analysis showed that TOX3 had no effect on FOXA1 protein or mRNA levels (data not shown). With respect to its expression in luminal cells (Fig. 16), TOX3 mirrors the expression pattern of FOXA1 more than that of other luminal transcription factors. For instance, GATA3 is expressed at comparable levels in both luminal progenitors and mature luminal cells, while FOXA1 is predominantly expressed in mature luminal cells (Fig. 16 C), thus correlating with ERα expression levels. This is in keeping with its function as a pioneer factor of ERα-dependent transcription. FOXA1 binds at distal enhancer elements by displacing histones and opening up the chromatin structure, thereby facilitating the recruitment of ERα and the transcriptional machinery at proximal target gene promoters (Carroll and Brown, 2006; Cirillo et al., 2002). It is tempting to speculate that TOX3 functions in a way that is similar to FOXA1 to enable ERα-dependent transcription. TOX3 binding to distal enhancer elements rather than in proximity of ERα response elements (ERE) would explain why it TOX3 had no effect on ERE-dependent transcription in the luciferase assay (Fig. 35) since the reporter constructs only contained isolated ERE or short proximal promoter elements.

The preliminary TOX3 microarray data showed a TOX3-dependent decrease in GATA3 expression (Fig. 33). Loss of GATA3 in early tumours is linked to a loss of differentiation and marks tumour progression in a GATA3-MMTV-PyMT tumour model (Kouros-Mehr et al., 2008), and SNPs in the GATA3 gene have been shown to be associated with an increased breast cancer susceptibility (Garcia-Closas et al.,
In the context of the luminal transcriptional network, it would certainly be of interest to further investigate a correlation between TOX3 and GATA3 expression.

### 4.2.3 TOX3 and ESR1

GSEA of the TOX3 microarray data suggested that TOX3 had, if any, a weak suppressive effect on ESR1 expression itself (Fig. 33 I), and Western blot analysis confirmed this finding (Fig. 34). Since ESR1 expression is not detectable in unsorted primary mammary epithelial cells, it was not surprising that no difference in ESR1 levels was detected in TOX3 versus control transduced BPEC (Fig. 43 A). In keeping with the microarray data, separate overexpression of TOX3 and ESR1 in BPEC did not result in the same cell morphology, indicating that the regulation of luminal differentiation by TOX3 does not depend on ESR1 and vice versa, or at least not to a great extent (Fig. 43 B). This is not surprising, as it is not expression of ESR1 that ultimately defines the luminal phenotype, a notion that is exemplified by the absence of ESR1 expression in a subset of normal mature luminal cells. Moreover, ESR1 expression is restriction to the luminal A and B breast cancer subtypes, while other luminal differentiation genes are expressed in the “broad luminal” subgroup that also includes molecular apocrine tumours, which are ERα-negative tumours that are characterised by androgen signalling and do not classify as basal (Farmer et al., 2005). In keeping with a role for TOX3 in a more widely defined luminal cell differentiation, TOX3 also repressed the molecular apocrine gene activated leukocyte adhesion molecule (ALCAM) (Fig. 33 I).

GSEA further showed that TOX3 expression resulted in downregulation of the classic ESR1 target gene growth regulation by estrogen in breast cancer 1 (GREB1) (Fig. 33 I), expression of which has been shown to closely reflect S phase entry of MCF-7 cells upon stimulation by beta-estradiol (Deschenes et al., 2007; Rae et al., 2005). In contrast, another classic ESR1 target gene, TFF1, was not suppressed by TOX3 (Fig. 33 I), which is consistent with TOX3 having no transcriptional effect on the TFF1 promoter (Fig. 35).

As discussed below, due to the quality of the data, the relevance of the TOX3 microarray results is unclear at this point.
4.2.4 TOX3 and CITED1

To date, only two studies have been published that characterise TOX3 function at the molecular level, both of which are concerned with the role of TOX3 in neurons. Using deletion mapping, these studies have shown that TOX3 contains an N-terminal transactivation domain that regulates transcription in concert with the HMG-box, while the C-terminal domain interacts with TOX3 binding partners CREB and CITED1 (Dittmer et al., 2010; Yuan et al., 2009). CITED1 is of particular interest because it has been implicated in ERα signalling (Kim et al., 2011; McBryan et al., 2007; Yahata et al., 2001) and has been shown to act as a transcriptional co-activator in numerous hormone-regulated tissues, including the mouse mammary gland, pituitary and ovaries (Kim et al., 2011; McBryan et al., 2007; Sriraman et al., 2010). Moreover, the Visvader study ranked both TOX3 and CITED1 among the ten genes that show the highest conservation between human and mouse in the mature luminal subpopulation (Table 8), and the expression pattern of CITED1 resembles that of TOX3 in the mammary epithelial subpopulations (Fig. 16) (Lim et al., 2010).

It is intriguing to speculate that both proteins also interact in mature luminal cells to activate a luminal-specific transcriptional programme. However, a potential interaction of TOX3 and CITED1 in mammary cells is purely speculative at this point. Using co-immunoprecipitation, I am currently investigating if TOX3 interacts with CITED1 in mammary cells. If the outcome is positive, it will be of interest to examine whether TOX3 is dependent on cooperation with CITED1 to activate transcription of ERE-containing promoters, as observed by Dittmer and colleagues. I observed that TOX3 alone was not able to act as a transcriptional activator of ERE-containing minimal promoter elements (Fig. 35). It is unclear why my luciferase data contradict the observations made by Dittmer and co-workers, although it has to be noted that the latter reported an increase of target promoter activity of approximately fivefold (Dittmer et al., 2010), which appears rather small. Differences in TOX3 activity may further be due to the use of different cell types, a lack of required co-activators such as CITED1, which facilitate binding to the DNA binding site or mediate interaction with a third transcriptional regulator such as ESR1. It appears unlikely that CITED1 alone is capable of recruiting of TOX3 to target promoters because it lacks a DNA-binding domain. Tandem affinity purification did not yield any information about a possible interaction between TOX3 and CITED1 since
CITED1 is not expressed in HEK 293T cells (data not shown). It remains unclear how TOX3 regulates luminal mammary epithelial cell differentiation without being involved in estrogen-dependent signalling but treatment of MCF-7 cells with beta-estradiol did not change TOX3-dependent transcriptional regulation of ERE-containing promoters (Fig. 35), nor did it have an influence on TOX3 expression levels themselves (data not shown). This was consistent with the findings by Dittmer and colleagues, who showed that treatment with fulvestrant did not inhibit TOX3-dependent transcriptional activation of ERE-containing promoters (Dittmer et al., 2010). The significance of this finding is unclear, as both the work presented here and the latter study examined a limited number of artificial promoter constructs. Regulation of ERα target genes by TOX3 may be more complex. The microarray analysis of TOX3 target gene expression in MCF-7 cells supports this notion, as some ERα target genes such as PGR were repressed in the presence of TOX3 while others such as TFF1 were not inhibited (Fig. 33 I). Similarly, transcriptional co-regulation of ERα target genes by CITED1 appears to be gene-specific, as in MCF-7 cells, CITED1 activates transcription of TGF-α in an E2-dependent manner but not of TFF1 (Yahata et al., 2001). Microarray analysis of untreated versus E2-treated TOX3 expressing cells may be useful to shed light on estrogen-dependence of TOX3 transcriptional regulation. It is of note that I did not find any effect of TOX3 on CITED1 expression (data not shown) but the microarray data suggested that TOX3 expression resulted in downregulation of the related transcription factor CITED2 (Fig. 33). No data regarding a potential interplay between CITED1 and CITED2 in the breast has been reported to date but it has been shown that they are differentially expressed in the developing kidney (Boyle et al., 2007).

4.2.5 TOX3 potentially interacts with YBX1

Using tandem affinity purification in combination mass spectrometric analysis of TOX3, I have identified the transcription/translation factor Y-box binding protein 1 (YBX1, YB-1) as a potential TOX3 interacting protein. However, since the mass spectrometry data were of low quality due to high keratin contamination, only very abundant proteins were detected in the NTAP-TOX3 sample, such as chaperone proteins cytoskeletal proteins. The expression data from the Visvader group appear to
confirm that YBX1 is an abundant protein in all mammary subpopulations (Fig. 16 C) (Lim et al., 2010). Therefore, the mass spectrometry results need to be considered with some caution, and the interaction between TOX3 and YBX1 needs to be confirmed by co-immunoprecipitation. In addition, since the NTAP control sample was not analysed by mass spectrometry because of the keratin contamination, the possibility cannot be excluded that YBX1 in fact bound to the TAP tag and not to TOX3. Nevertheless, YBX1 would be a potentially interesting candidate as it is known to act as a breast cancer oncogene that promotes tumour growth through a multitude of functions including the induction of genes such as EGFR, ERBB2, PCNA, PIK3CA and ATP-binding cassette, sub-family B (MDR/TAP), member 1/multidrug resistance protein 1 (ABCB1/MDR1) (Astanehe et al., 2009; Bargou et al., 1997; Ise et al., 1999; Jurchott et al., 2003; Stratford et al., 2007; Wu et al., 2006). YBX1 is phosphorylated at serine residue 102 in response to mitogenic stimulation, which is thought to be required for its transcriptional activity (Stratford et al., 2007). A recent study showed that YBX1 was phosphorylated in response to KRAS activation, which in turn was induced by ERBB1 signalling or ionising radiation and was mediated through the PI3K or MAPK1 axes (Toulany et al., 2011). Using chromatin immunoprecipitation on chip (ChIP-on-chip) analysis, Dunn and co-workers identified the met proto-oncogene (hepatocyte growth factor receptor) (MET), CD49f and CD44 as well as several members of the NOTCH and WNT pathways as YBX1 transcriptional target genes (Finkbeiner et al., 2009). In a follow-up study they showed that YBX1 enhances self-renewal and promoters growth as well as drug resistance of basal breast cancer cells through induction of the CD49f and CD44 genes (To et al., 2010).

I am currently carrying out co-immunoprecipitation of HA epitope-tagged TOX3 and YBX1 in MCF-7 and BT-474 cells to confirm the association of TOX3 and YBX1 in a luminal environment. Several studies have suggested that YBX1 plays a role in basal breast cancer cells in particular (Finkbeiner et al., 2009). It would therefore also of interest to investigate whether TOX3 physically interacts with YBX1 in a basal breast cancer line such as MDA-MB-231 in a manner that is different to the luminal cell lines. What function a potential interaction between TOX3 and YBX1 could have remains to be determined but since both act as transcriptional activators, it appears likely that they co-regulate a set of target genes. Preliminary comparison of my TOX3 microarray data and YBX1 target genes did not reveal any obvious genes that were
regulated by both transcription factors (Finkbeiner et al., 2009), although both proteins appear to have a pro-proliferative function. Alternatively, one could speculate that TOX3 inhibits YBX1 by binding to it, thereby preventing YBX1-dependent upregulation of CD49f and CD44 and self-renewal of normal progenitor cells.

As discussed in the Results section and above, keratin contamination and association of misfolded TOX3 with chaperone proteins likely prevented the identification of further TOX3 interacting proteins by mass spectrometric analysis of tandem affinity purified TOX3. However, the proteomic approach promises to be a valuable technique to identify TOX3 interacting proteins. In recent years, the Agilent InterPlay Mammalian tandem affinity purification (TAP) protocol has been successfully used to identify protein-protein interactions in a number of studies, including interactions between chromatin-binding polycomb proteins BMI1 and polycomb group ring finger 2 (PCGF2) as well as the antiviral apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G (APOBEC3G) and YBX1 (Gallois-Montbrun et al., 2007; Wiederschain et al., 2007). It is of note that in the latter study, YBX1 bound to APOBEC3G in ribonucleoprotein (RNP) complexes and stress granules and that the interaction could be abolished by RNase treatment (Gallois-Montbrun et al., 2007). The significance of the RNA-binding properties of YBX1 for a potential interaction with TOX3 is unclear but should be considered during further analysis. To identify further proteins that TOX3 forms a complex with, I am currently repeating the tandem affinity purification using lentivirally transduced HEK 293T cells. Separation of the purified proteins by SDS-PAGE and subsequent colloidal blue staining of the gel will be carried out in a clean room to avoid keratin contamination as much as possible. While HEK 293T cells are commonly used for TAP experiments (Gallois-Montbrun et al., 2007), they are not suited for the identification of mammary cell-specific TOX3-containing complexes. Therefore, I plan to carry out TAP using MCF-7 cells transduced with the TAP-TOX3 constructs as the luminal MCF-7 breast cancer cell line may provide a more physiological setting for the identification of TOX3 protein-protein interactions. For the same reason, TOX3-containing protein complexes would ideally be purified from cells that express endogenous TOX3, such as the BT-474 cell line. I have created BT-474 cells that stably express TAP-TOX3 (data not shown) but BT-474 cells have proved to proliferate slowly and the large number of cells required for TAP and subsequent mass spectrometric analysis are limiting factors for the use of
Discussion

this cell line. Similarly, although I initially set out to use primary mammary epithelial cells for TAP, it simply does not appear to be feasible to use BPEC transduced with TAP-TOX3 for the protein purification owing to the large numbers of cells required.

4.3 A role for TOX3 in luminal breast cancer

4.3.1 TOX3 is expressed in luminal breast tumours

Immunohistochemical (IHC) staining of primary breast tumour sections showed TOX3 expression in luminal but not in basal-like breast tumours (Fig. 12), which confirmed the tumour expression array data (Fig. 11). In contrast, IHC staining of normal tissue adjacent to tumour tissue in the tumour microarray biopsies further detected TOX3 in both luminal and myoepithelial cells (Fig. 12), which was at odds with the available transcriptional profiling data of myoepithelial cells (Fig. 16). A possible explanation is the lack of purity of the TOX3 antibodies used for immunohistochemistry. Western blot analysis showed that the custom-made rabbit polyclonal TOX3 antibody successfully detected ectopically expressed TOX3 (Fig. 28, 29). However, high non-specific background staining complicated the analysis of endogenous TOX3 protein expressed at lower levels (Fig. 15, 29). The unavailability of a high-quality TOX3 antibody hampered the analysis of TOX3 protein expression throughout this project and has to be addressed in future investigations. To determine whether TOX3 is a good prognostic marker that can be used for IHC staining in the clinical setting, the availability of a high-quality TOX3 antibody is of particular interest.

4.3.2 TOX3 induces proliferation and represses luminal differentiation genes in MCF-7 cells – a tentative interpretation

The above data leave open the question of how TOX3 contribute to the formation of ERα-dependent breast cancer. Gene set enrichment analysis of the TOX3 microarray showed that TOX3 expression led to the downregulation of genes that are usually overexpressed in ERα-positive breast cancer such as GREB1 and GATA3, as described above. In addition, TOX3 expression in MCF-7 cells resulted in the
induction of proliferation genes such as MYC, CCNA2, PCNA, PLK1 as well as inhibition of CDKN1A (Fig. 33). However, the low quality of the expression data must be considered when interpreting the results. Not only did the induction of interferon response genes in the control samples complicate the interpretability of the results (Fig. 33) but the proliferation signature itself should be considered with caution as it may have been caused by inadvertent interference with cell proliferation and subsequent release of the cells into the cell cycle caused by the lentiviral infection.

Taking into account these caveats, based on preliminary interpretation of the microarray data, it is tempting to speculate that TOX3 plays a role in the development of ERα-positive breast cancer by attenuating luminal cell differentiation through inhibition of a subset of ERα target genes and by simultaneously inducing proliferation (Fig. 49). Inhibition of differentiation genes would allow ERα-positive cells to switch from a paracrine to an autocrine mode with respect to growth-promoting signalling, thus enabling them to directly respond to estrogen stimulation. It remains to be determined how this can be reconciled with a role for TOX3 in the differentiation of normal luminal epithelial cells. Tight spacial and temporal control of TOX3 expression during luminal cell development may be a requirement for its appropriate function. TOX3 expression levels in turn may be influenced by the polymorphisms in the regulatory region of the gene that have been associated with an increased breast cancer risk.

As in vitro and in vivo analysis of TOX3 expression in MCF-7 cells showed, overexpression of TOX3 was associated with an initial transient growth advantage both in culture and in xenografts but this proliferative advantage could not be maintained over a longer period of time (Fig. 45, 46), suggesting that TOX3 may exert a potential oncogenic effect in a manner that is more complex than a simple induction of proliferation. One possible explanation may be that TOX3 facilitates the adhesion or invasiveness of TOX3 expressing cells, which promote initial engraftment of transformed cells in the neighbouring normal tissue, as seen in the intraductal xenograft model (Fig. 46). The TOX3 microarray data suggest an upregulation of both the chemokine (CXC motif) ligand 14 (CXCL14) and the chemokine (C-X-C motif) receptor 4 (CXCR4) in TOX3-expressing cells. The Polyak group has shown that CXCL14 promotes invasion, migration and proliferation, while
recent studies have demonstrated that CXCR4 mediates E2-induced proliferation of ERα-positive tumour cells and is capable of transforming MCF-10A cells (Allinen et al., 2004; Boudot et al., 2011; Su et al., 2011). Furthermore, microarray analysis of ESR1 expression in normal HMEC previously conducted in our lab revealed CXCL14 as one of the genes that was most strongly induced by ESR1 expression in a ligand-independent manner (S. Duss, personal communication). Again, based on the poor expression data, the observed induction may be an artefact but it may be of interest to further investigate the relationship between TOX3 and these mediators of chemokine signalling. It is of note however, that TOX3 expression also resulted in a significant downregulation of AGR2 expression (Fig. 33, 34). AGR2 has been shown to promote tumour growth and metastasis (Brychtova et al., 2011). The implications of the seemingly opposing roles of TOX3 in repression and induction of oncogenic genes remain to be determined.

In the context of mammary epithelial cell regulation, a potentially interesting result was that TOX3 expression resulted in a strong repression of progesterone receptor (PGR) expression in MCF-7 cells (Fig. 33). PGR as well as insulin receptor substrate 1 (IRS1) were in fact among the most downregulated genes. Like aforementioned luminal genes, PGR and IRS1 are ESR1 target genes. In addition, both are involved in a complex regulatory relationship with ESR1 in mammary cells (Chan and Lee, 2008; Conneely et al., 2003).

IRS-1 is a signal transduction protein that binds directly to insulin-like growth factor 1 receptor (IGFR1) and insulin receptor (IR) in the canonical signalling pathway. Downstream effectors of IRS-1 and the structurally and functionally closely related IRS-2 include phosphoinositide-3-kinase (PI3K) and mitogen-activated protein kinases 1 and 3 (MAPK1/3) in a multitude of tissues including the mammary gland. IRS1 functions in mammary gland development, and in keeping with the function as a transducer of mitogenic signals, IRS1 has been shown to be a transforming oncogene in the breast (Chan and Lee, 2008). Interestingly, in mice, loss of IRS1 leads to an increase in metastasis due to compensation by IRS2-mediated signalling (Ma et al., 2006), which would be consistent with the model of TOX3 function that I propose (Fig. 49). However, in MCF-7 cells downregulation of IRS1 has been shown cause apoptosis (Nolan et al., 1997). It remains to be determined how these opposing findings can be reconciled and if and how TOX3 regulates IRS1 expression, as Western blot analysis did not confirm repression of IRS1 by TOX3 (Fig. 34).
In contrast, Western blot analysis did confirm that TOX3 repressed both isoforms PGR-A and PGR-B in MCF-7 and T-47D cells (Fig. 34 and data not shown). Knockdown of endogenous TOX3 does not appear to upregulate PGR levels in BT-474 and MDA-MB-361 cells, which may be explained by insufficient knockdown efficiency of the shRNAmir constructs (Fig. 30). An alternative explanation may be differences in the regulation of PGR expression in TOX3-deficient MCF-7 cells and BT-474 or MDA-MB-361 cells. As discussed above, the functional relationship between TOX3 and ESR1 remains unclear but it is possible that the effect of TOX3 on PGR expression is independent of ESR1. Despite a degree of cross-talk between estrogen and progesterone signalling, notably the induction of PGR expression by estrogen at the onset of puberty (“estrogen priming”) (Haslam and Shyamala, 1979), ESR1-mediated control of mammary gland development predominates during puberty, while PGR governs side-branching during estrous cycles and, in concert with the prolactin receptor (PRLR), pregnancy (Brisken and O’Malley, 2010). Reconstitution studies in mice have shown that PGR−/− cells are capable of generating chimeric mammary glands when mixed with wild-type cells, demonstrating that the mitogenic stimulus that leads to the proliferation of MaSC and early progenitor cells is exerted in a paracrine manner (Brisken et al., 1998). As discussed above, RANKL and WNT4 have been identified as mediators of progesterone signalling and MaSC activation during the estrous cycle (Asselin-Labat et al., 2010; Brisken et al., 2000; Joshi et al., 2010). Progesterone stimulation induces proliferation in neighbouring cells at least in part by activating expression of cyclin D1 (CCND1) (Sicinski et al., 1995) but the exact mechanism of paracrine stimulation of proliferation including the involvement of other transcription factors acting upstream of PGR are poorly understood.

Considering the pro-proliferative role of PGR in mammary gland development and MaSC activation, it is difficult to conceive how TOX3 would promote oncogenic transformation by downregulating PGR expression in luminal cells. In addition, the majority of luminal tumours are PGR-positive, which is at odds with TOX3-dependent repression of PGR.
4.3.3 Creation of a luminal tumour model by quantitative transformation of normal human mammary epithelial cells

As discussed above, the possibility remains that the TOX3 microarray results are simply an experimental artefact and that the observed repression of luminal differentiation and induction of proliferation were not caused by TOX3 expression. Alternatively, MCF-7 cells may be entirely unsuited as a model system and expression of the potential luminal differentiation factor TOX3 in MCF-7 cells does not reproduce what happens in vivo in normal human cells or in human breast cancer.

To test if TOX3 is a transforming oncogene, it is therefore preferable to express TOX3 in normal human mammary epithelial cells. To this end, I have successfully established a culture system that combines the initial enrichment for progenitor cells by mammosphere suspension culture with subsequent adherent culture in WIT medium on Primaria plastic, which allows longer maintenance of primary mammary epithelial cells (BPEC) in culture than conventional cell culture media (Fig. 41). Despite improved longevity of primary human mammary cells in culture owing to increasingly sophisticated culture methods, it remains preferable to limit the in vitro culture as much as possible. Even when using optimised culture conditions, it is evident that the cellular phenotype changes over the course of time and a certain degree of differentiation cannot be avoided. For instance, despite their improved in vitro life span, BPEC grown in WIT medium exhibited a secretory phenotype after the first passages, which was characterised by the appearance of large vacuole-like inclusions in the cytoplasm (data not shown). On the other hand, this phenotype was not always detectable, suggesting a certain variability of the WIT medium-based cell culture system. In addition, culture of organoid-derived BPEC in WIT medium resulted in progressive disappearance of cells of luminal epithelial differentiation (Fig. 40), which is a feature of HMEC cultured in HMM+ medium but which is not consistent with the report by Ince and colleagues (Ince et al., 2007). In my hands, culture in WIT medium therefore was not able to maintain the different mammary epithelial subpopulations but favoured myoepithelial differentiation. In addition, BPEC invariably underwent growth arrest and displayed a senescent phenotype, which was in contrast to the claim made by Ince and co-workers that culture in WIT medium allowed theoretically infinitive maintenance of BPEC (Ince et al., 2007). A further difference to the culture system described by the Weinberg lab was that BPEC
transformed with TERT, MYC, BMI1 and ESR1 (BPEC XS08.4) failed to form adenocarcinomas in NSG mice but gave rise to tumours of squamous differentiation, much like transformed HMEC (Fig. 48), while BPEC transformed with TERT, SV40 large T antigen and oncogenic HRAS-V12 were reported to form adenocarcinomas in the xenograft model (Ince et al., 2007; McAllister et al., 2008). Squamous differentiation of BPEC XS08.4-derived tumours was likely due to inadequate differentiation of the cells while in culture or to the expression of a combination of genes that do not exactly mirror the set of genes that are overexpressed in human luminal tumours, or both. Our group is currently testing different combinations of transgenes in order to prevent squamous differentiation of BPEC-derived tumours.

Considering the role of GATA3 in luminal differentiation, it is an interesting candidate transgene to test. Furthermore, the next step will be to express and characterise TOX3 in the BPEC luminal tumour model.

In my opinion, it is crucial to shorten the in vitro cell culture time as much as possible in order to avoid cell culture artefacts. With respect to the establishment of a luminal breast cancer model by lentiviral transduction of normal cells, this means that it is indispensable to optimise lentiviral expression systems in order to shorten the time required for serial transduction and selection. One limitation of the lentiviral system presented here is that all transgenes were expressed from the same lentiviral backbone, which complicates the selection of successfully transduced cells, which was seen in the BPEC XS.08 line (Fig. 48). To ensure simultaneous transduction with multiple viruses, the use of a combination of different fluorescent and antibiotic resistance markers that allow for visual, flow cytometric and antibiotic selection would be ideal. Simultaneous expression of multiple transgenes further requires the production of viral preparations at high titres. At the same time, overly high multiplicities of infection (MOI) must be avoided as they may trigger an acute interferon response in infected cells, as exemplified by the MCF-7 GUS control cells used in the microarray experiment (Fig. 33). Other possible causes for the induction of interferon responses include the contamination of plasmid preparation with endotoxin, which can be reduced by employing appropriate plasmid purification methods, and the innate immunity to viral intermediates such as RNA-DNA hybrids or misfolded RNAs. Wild-type lentiviruses have evolved to inactivate mediators of the host cell interferon response (Haller et al., 2006) but since the majority of viral genes have been deleted from lentiviral vectors for safety considerations, viral
intermediates most likely cause an interferon response in packaging cells at the time of virus production as well as in target cells at the time of infection. To avoid interferon response artefacts, it is therefore preferable to analyse the effect of transgene expression later than 48 hours following lentiviral infection to allow for the acute interferon response to subside. However, if the immediate effects of transgene expression are of interest, this may cause a problem. Albeit technically more challenging than constitutive vectors, the use of inducible lentiviral vector systems allows to separate the time of infection and the time of transgene expression. The KRAB repressor/tet-off conditional lentiviral system for TOX3 expression (Szulc et al., 2006) will hopefully also help to prevent silencing of TOX3 expression that was observed with the hPGK lentiviral construct. Alternatively, silencing of TOX3 may be tested in the context of other promoters, although the hPGK promoter had been originally been chosen for its constitutive expression across the cell types of interest, while expression from other promoters such as the CMV promoter appeared to vary considerably between cell types (Duss et al., 2007 and S. Duss, personal communication). Variegation due to epigenetic silencing is a problem frequently encountered with lentiviral expression systems (Golding et al., 2010; Zhang et al., 2010). One way to avoid silencing of transgenes by methylation, histone deacetylation and chromatin condensation is to introduce silencing-resistant elements into the promoter region of lentiviral expression vectors (Williams et al., 2005). However, what speaks against silencing of the hPGK promoter is that the same lentiviral backbone has been used successfully in our laboratory to express a number of transgenes without obvious silencing effects. This would suggest that the observed silencing was intrinsic to the TOX3 ORF and that TOX3 expression was subject to negative regulation, possibly because its expression was disadvantageous to the cell. The microarray data suggested that TOX3 itself did not repress PGK1 expression (Fig. 33 I). In addition, regulation of TOX3 expression did not appear to occur at the protein level as treatment with the proteasome inhibitor MG132 had no obvious effect on TOX3 protein levels (data not shown). The cause for TOX3 silencing remains to be determined but the problem shows that the choice of lentiviral expression system is of great importance.

In this study, two different xenograft methods were used. The intraductal injection method worked very well for the MCF-7 breast cancer cell line (Fig. 46) and arguably
represents the currently available method that most accurately replicates the site of tumour initiation in the human breast (Behbod et al., 2009). In contrast, unpublished data from our lab have shown that transformed BPEC are not able to proliferate in the ducts following intraductal injection, presumably because they are not invasive enough to replace murine epithelial cells in their niches, or because they lack the adhesion molecules that allow engraftment in the murine ducts. For the quantitative transformation BPEC model, I therefore chose the orthotopic injection into the forth mammary gland, which had proved successful for the assessment of the tumourigenic properties of HMEC transformed with BMI1, MYC, TERT and ESR1 in that tumours formed readily and cells were able to metastasise to distant organs (Duss et al., 2007). In keeping with results from the luminal HMEC model, quantitatively transformed BPEC injected into the fourth mammary gland efficiently gave rise to tumours (Fig. 48). The choice of xenograft model system has been subject of much recent research and commonly employed alternatives to the orthotopic xenotransplantation model include subcutaneous injections and, particularly for human mammary stem cells assays, injections under the renal capsule (Eirew et al., 2010). My rationale for choosing the orthotopic xenograft model was that, despite the aforementioned differences between humans and mice, the murine mammary gland resembles the human breast in its stromal microenvironment and ductal architecture more closely than other possible sites of injection. A second point of recent discussion has been the use of slow-release estrogen pellets to mimic serum estrogen levels of pre-menopausal women (Anderson et al., 1998). No obvious difference was observed between insertion of estrogen pellets one week before or at the time of xenografts (data not shown). However, HMEC transformed with TERT, BMI1, ESR1 and MYC failed to grow in the absence of exogenous estrogen (Duss et al., 2007). A recent study by the Vonderhaar group found no significant difference in tumour growth or latency between xenotransplants of patient-derived human breast cancer cells in the presence versus absence of E2 supplements (Fleming et al., 2010). This is in contrast to our findings but is likely due to the phenotype of the injected cells. Furthermore, I found that injection of an admixture of transformed BPEC (or HMEC) and human mammary fibroblasts as well as 12.5% matrigel gave the best results in terms of tumour incidence, suggesting that a certain degree of “humanisation” of the murine mammary fat pad benefits engraftment of human cells.
4.3.4 The predicted non-coding RNA gene LOC643714

The SNPs that are associated with an increased breast cancer risk are located in the upstream regulatory region of the TOX3 gene, which overlaps with the most 3’ exon of the predicted non-coding RNA gene LOC643714 (Easton et al., 2007; Ruiz-Narvaez et al., 2010; Stacey et al., 2007; Udler et al., 2010). The possibility therefore exists that the genetic variations associated with breast cancer predisposition are in fact linked to the LOC643714 gene and not to TOX3. To date, the LOC643714 gene remains uncharacterised but regulatory non-coding RNAs are known to affect the expression of their target genes in various ways, ranging from interference with cis-acting binding sites for regulatory proteins in the promoter regions of adjacent genes to regulating transcript stability. An example for the former is the regulation of X chromosome inactivation by Xist RNA, while the latter is exemplified by trans-acting miRNAs (Mattick, 2009). What speaks against a causative role for LOC643714 is that while it appears to be expressed in the human breast, its expression levels remain constant across the different mammary epithelial subpopulations as well as the stroma (Fig. 16 C), although the lack of variability could mean that the microarray probe gave no signal.

4.4 How does TOX3 cause breast cancer?

A central question of this study was whether SNPs in the TOX3 gene lead to a gain or a loss of TOX3 function. The MCF-7 GSEA results and the MCF-7 xenograft model suggest that TOX3 is an oncogene, which leaves open the question how a potential oncogene can be so closely linked to terminal differentiation. Figure 48 summarises the currently available data in a tentative model of TOX3 function in normal and luminal breast cancer cells. It remains to be seen whether TOX3 determines terminal differentiation or whether it is only part of a functioning differentiated luminal epithelial cell like genes such as AGR2 or XBP1.
Figure 49. Model of TOX3 function in normal and tumour cells

Schematic of potential TOX3 function and the factors regulated by TOX3 in the normal mammary epithelium and luminal breast cancer. Potential TOX3 target genes are in blue, while potential TOX3 interacting proteins are in black.

4.5 Concluding remarks and future perspectives

I here present evidence that the nuclear HMG-box protein TOX3 is involved in the regulation of both mammary epithelial cell fate in the normal breast and the oncogenic transformation of luminal epithelial cells. This work could only be an initial characterisation of TOX3 function but it will hopefully serve as a starting point for further investigations.

Future work will focus on the characterisation of TOX3 function by lentivirus-based expression in normal human mammary epithelial cells. Comparison of the effect of ectopic expression of TOX3 in luminal tumour cell lines such as MCF-7 and the bipotent immortalised cell line MCF-10A will permit to examine TOX3 function in a variety of environments. It will be of great interest to repeat the microarray analysis of TOX3 target genes in normal human mammary epithelial cells to identify differences
in TOX3 function in normal versus cancerous cells. The inducible lentiviral TOX3 vector will permit to better control TOX3 expression. Transcriptional profiling of TOX3-expressing MCF-7 cells has already provided a list of candidate target genes, the most interesting of which will be subject of further investigation. The oncogenic potential of TOX3 is currently being tested in the quantitative transformation protocol for BPEC that I have established. Knockdown of endogenous TOX3 in the luminal breast cancer lines will provide further insight into TOX3 function. Characterisation of the transcriptional activity of TOX3 is of major interest. Genome-wide mapping of TOX3 binding sites using chromatin immunoprecipitation (ChIP) analysis requires reliable antibodies, which is currently a limiting factor in the analysis of TOX3 protein. An alternative to classic ChIP analysis using TOX3-specific antibodies involves in vivo biotinylation of TOX3 and subsequent affinity capture using a streptavidin-coupled matrix (Kim et al., 2008; Wang et al., 2006). Forced TOX3 expression in a transgenic mouse model as well as TOX3 knockout mice will ultimately show whether TOX3 has a function in development of the mammary gland. Transcriptional regulators are not prime drug targets for the treatment of breast cancer but if a role for TOX3 in breast cancer development can be confirmed, TOX3 expression in luminal tumours and the possibility to test for inherited genetic variations in the breast cancer susceptibility gene TOX3 may have prognostic value. Furthermore, identification of the signalling pathways that TOX3 is involved in may contribute to tailoring therapeutic approaches.
5 References


References


188


Vandesompele, J., De Preter, K., Pattyn, F., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3, 34.31-34.11.


# Appendix

## 6.1 Plasmid vectors used in this study

Table 11. Plasmid vectors used in this study.

<table>
<thead>
<tr>
<th>pXS plasmid name</th>
<th>Original plasmid name</th>
<th>Description</th>
<th>Supplier/cloned by</th>
</tr>
</thead>
<tbody>
<tr>
<td>pXS-3</td>
<td>pRLLsin.PPTs.hCMV.GFPpre</td>
<td>CMV-GFP pRRL viral expression vector</td>
<td>L. Naldini</td>
</tr>
<tr>
<td>pXS-10</td>
<td>Ultimate ORF clone IOH61112</td>
<td>TOX3-pENTR221 Gateway entry vector</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pXS-11</td>
<td>pMD2.G</td>
<td>VSV-G envelope protein construct for lentivirus pseudotyping</td>
<td>L. Naldini</td>
</tr>
<tr>
<td>pXS-16</td>
<td>PCMVdR8.74</td>
<td>Lentivirus gagpol packaging construct</td>
<td>D. Trono</td>
</tr>
<tr>
<td>pXS-17</td>
<td>pRRLsin.PPTs.hPGKGFPpre</td>
<td>GFP under control of hPGK promoter</td>
<td>L. Naldini</td>
</tr>
<tr>
<td>pXS-32</td>
<td>pNTAP-A</td>
<td>Interplay Mammalian TAP system parental cloning vector</td>
<td>Agilent Technologies</td>
</tr>
<tr>
<td>pXS-35</td>
<td>pDONR201</td>
<td>Gateway acceptor entry vector</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pXS-36</td>
<td>pXS-36</td>
<td>hPGK-TOX3-mPGK-puro pRRL viral vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-37</td>
<td>pSD-104</td>
<td>pENTR-DsRed2</td>
<td>S. Duss</td>
</tr>
<tr>
<td>pXS-38</td>
<td>pJH-3982</td>
<td>cPPT-hUbiC-attR-hPGK-EGFP-WPRE-pRRL destination lentiviral vector</td>
<td>J. Huelsken</td>
</tr>
<tr>
<td>pXS-39</td>
<td>pXS-39</td>
<td>cPPT-hUbiC-TOX3-hPGK-EGFP-WPRE-pRRL lentiviral vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-41</td>
<td>pXS-41</td>
<td>cPPT-hUbiC-TOX3-hPGK-EGFP-WPRE-pRRL lentiviral vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-48</td>
<td>pJH-4282</td>
<td>cPPT-CMV-attR-mPGK-GFP-WPRE-pRRL dest lentiviral vector</td>
<td>J. Huelsken</td>
</tr>
<tr>
<td>pXS-62</td>
<td>Clone ID RHS4346</td>
<td>pGIPZ shRNAmir scrambled negative control lentiviral vector</td>
<td>Open Biosystems</td>
</tr>
<tr>
<td>pXS-70</td>
<td>pXS-70</td>
<td>TOX3-GFP Gateway entry vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-71</td>
<td>pXS-71</td>
<td>hPGK-TOX3-GFP-mPGK-puro pRRL lentiviral vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-72</td>
<td>pXS-72</td>
<td>cPPT-CMV-TAP-mPGK-GFP-WPRE-pRRL lentiviral vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-85</td>
<td>pXS-85</td>
<td>cPPT-CMV-NTAP-TOX3-mPGK-GFP-WPRE-pRRL lentiviral vector</td>
<td>X. Schmidt, E. Richard</td>
</tr>
<tr>
<td>pXS-88</td>
<td>Clone ID RHS4430-98480880 - V2LHS_75758</td>
<td>pGIPZ shRNAmir lentiviral vector against TOX3 #1 – Target sequence: GACATCTGATGACTATAA</td>
<td>Open Biosystems</td>
</tr>
<tr>
<td>pXS-89</td>
<td>Clone ID RHS4430-99157247 - V2LHS_75830</td>
<td>pGIPZ shRNAmir lentiviral vector against TOX3 #2 – Target sequence: GCCCTCTGTCATAGAA</td>
<td>Open Biosystems</td>
</tr>
<tr>
<td>pXS-93</td>
<td>pXS-93</td>
<td>NTAP-TOX3 Gateway entry vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>-------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>pXS-95</td>
<td>pXS-95</td>
<td>hPGK-NTAP-TOX3-mPGK-puro pRRL lentiviral vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-96</td>
<td>pXS-96</td>
<td>cPPT-hUbiC-TAP-TOX3-hPGK-EGFP-WPRE-pRRL lentiviral vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-97</td>
<td>pXS-97</td>
<td>TAP control Gateway entry vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-98</td>
<td>pXS-98</td>
<td>hPGK-NTAP-mPGK-puro pRRL lentiviral vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-99</td>
<td>pXS-99</td>
<td>cPPT-hUbiC-TAP-TOX3-hPGK-EGFP-WPRE-pRRL lentiviral vector</td>
<td>X. Schmidt</td>
</tr>
<tr>
<td>pXS-102</td>
<td>pGL3-PS2-luc</td>
<td>proximal PS2 promoter Luc reporter in pGL3-basic</td>
<td>G. Reid</td>
</tr>
<tr>
<td>pXS-103</td>
<td>ERE-TK-luc</td>
<td>Vitelloprotein ERE-5’ minimal TK (TATA) reporter Luc</td>
<td>G. Reid</td>
</tr>
<tr>
<td>pXS-104</td>
<td>pGL3-basic</td>
<td>pGL3-basic</td>
<td>Promega</td>
</tr>
<tr>
<td>pXS-105</td>
<td>pRL-CMV</td>
<td>pRL-CMV</td>
<td>Promega</td>
</tr>
<tr>
<td>pXS-106</td>
<td>pCF-332</td>
<td>pRL-SV40</td>
<td>Promega</td>
</tr>
<tr>
<td>pXS-107</td>
<td>pCF-333</td>
<td>pRL-TK</td>
<td>Promega</td>
</tr>
<tr>
<td>pXS-108</td>
<td>pEGFP-N1</td>
<td>CMV-MCS-EGFP</td>
<td>Clontech</td>
</tr>
<tr>
<td>pSD-69</td>
<td>pSD-69</td>
<td>hPGK-attR-mPGK-puro pRRL lentiviral destination vector</td>
<td>S. Duss</td>
</tr>
<tr>
<td>pSD-82</td>
<td>pSD-82</td>
<td>hPGK-ESR1-mPGK-puro pRRL lentiviral vector</td>
<td>S. Duss</td>
</tr>
<tr>
<td>pSD-83</td>
<td>pSD-83</td>
<td>hPGK-TERT-mPGK-puro pRRL lentiviral vector</td>
<td>S. Duss</td>
</tr>
<tr>
<td>pSD-84</td>
<td>pSD-84</td>
<td>hPGK-BMI1-mPGK-puro pRRL lentiviral vector</td>
<td>S. Duss</td>
</tr>
<tr>
<td>pSD-94</td>
<td>pSD-94</td>
<td>hPGK-MYC-mPGK-puro pRRL lentiviral vector</td>
<td>S. Duss</td>
</tr>
<tr>
<td>pER-6</td>
<td>pRL.sin.cPPT.MND.LUC.wpre</td>
<td>MND-luciferase lentiviral vector</td>
<td>E. Richard</td>
</tr>
<tr>
<td>pER-15</td>
<td>pER-15</td>
<td>hPGK-GFP-mPGK-puro pRRL lentiviral vector</td>
<td>E. Richard</td>
</tr>
<tr>
<td>pER-47</td>
<td>pER-47</td>
<td>hPGK-HA-TOX3-mPGK-puro pRRL lentiviral vector</td>
<td>E. Richard</td>
</tr>
</tbody>
</table>
Figure 50. Maps of lentiviral vectors used for tandem affinity purification.
### 6.2 Antibodies used in this study

Table 12. Antibodies used in this study.

<table>
<thead>
<tr>
<th>Antigen (Clone/epitope)</th>
<th>Supplier, Catalogue number</th>
<th>Raised in</th>
<th>Application (Dilution)</th>
<th>MW of antigen (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGR2</td>
<td>Abnova, Millipore 05-637</td>
<td>mouse mc</td>
<td>WB (1:1000)</td>
<td>20</td>
</tr>
<tr>
<td>BM11 (F6)</td>
<td>Abnova, Millipore 07-482</td>
<td>mouse mc</td>
<td>WB (1:1000)</td>
<td>40-44</td>
</tr>
<tr>
<td>CBP epitope tag</td>
<td>Abcam, ab15096 Lab Vision</td>
<td>rabbit pc</td>
<td>WB (1:2000) IF</td>
<td>depend. on rec. protein</td>
</tr>
<tr>
<td>CITED1</td>
<td>Thermo Fisher Scientific</td>
<td>rabbit pc</td>
<td>WB (1:250)</td>
<td>21</td>
</tr>
<tr>
<td>ESR1 (Sp1)</td>
<td>Thermo Fisher Scientific</td>
<td>rabbit mc</td>
<td>WB (1:1000) IF</td>
<td>67</td>
</tr>
<tr>
<td>ESR1 (1D5)</td>
<td>Dako, Abcam ab290</td>
<td>mouse mc</td>
<td>IHC</td>
<td>67</td>
</tr>
<tr>
<td>GFP (B-2)</td>
<td>Santa Cruz sc-9996</td>
<td>rabbit pc</td>
<td>WB (1:1000)</td>
<td>27</td>
</tr>
<tr>
<td>HA probe (F-7)</td>
<td>Santa Cruz sc-7392</td>
<td>mouse mc</td>
<td>WB (1:250)</td>
<td>depend. on rec. protein</td>
</tr>
<tr>
<td>phospho-H2AX (IBW301)</td>
<td>Upstate</td>
<td>mouse mc</td>
<td>IF (1:200)</td>
<td>15</td>
</tr>
<tr>
<td>HSP70 (SPA-810)</td>
<td>Stressgen, Enzo Life Sciences Millipore 05-1085</td>
<td>mouse mc</td>
<td>WB (1:1000)</td>
<td>70</td>
</tr>
<tr>
<td>IRS1 (4.2.2)</td>
<td>Abcam, Abisolutions sc-397</td>
<td>rabbit pc</td>
<td>WB (1:200)</td>
<td>185</td>
</tr>
<tr>
<td>KRT14 (LL001)</td>
<td>Birgit Lane</td>
<td>mouse mc</td>
<td>IF (1:5)</td>
<td>50</td>
</tr>
<tr>
<td>KRT14</td>
<td>Dako, Abcam ab290</td>
<td>mouse mc</td>
<td>IHC</td>
<td>50</td>
</tr>
<tr>
<td>KRT18 (LDK18)</td>
<td>Birgit Lane</td>
<td>mouse mc</td>
<td>IF (undiluted)</td>
<td>50</td>
</tr>
<tr>
<td>KRT18</td>
<td>Dako</td>
<td>mouse mc</td>
<td>IHC</td>
<td>50</td>
</tr>
<tr>
<td>MYC (9E10)</td>
<td>Santa Cruz sc-40</td>
<td>mouse mc</td>
<td>WB (1:500)</td>
<td>67</td>
</tr>
<tr>
<td>p16</td>
<td>Novocastra NCL-p16</td>
<td>mouse mc</td>
<td>WB (1:200)</td>
<td>16</td>
</tr>
<tr>
<td>p21 (C-19)</td>
<td>Santa Cruz sc-397</td>
<td>rabbit pc</td>
<td>WB (1:200)</td>
<td>21</td>
</tr>
<tr>
<td>pS3 (DO1)</td>
<td>David Lane</td>
<td>mouse mc</td>
<td>WB (1:1000)</td>
<td>53</td>
</tr>
<tr>
<td>PGR (A/B) (636)</td>
<td>Rockland Immunochemicals 600-401-252</td>
<td>mouse mc</td>
<td>WB (1:200)</td>
<td>81/116</td>
</tr>
<tr>
<td>TERT</td>
<td>rabbit pc</td>
<td>WB (1:500)</td>
<td></td>
<td>127</td>
</tr>
<tr>
<td>TOX3 Rb57</td>
<td>rabbit pc</td>
<td>WB (1:1000), IHC, IP</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>TOX3 Rb57P</td>
<td>rabbit pc</td>
<td>WB (1:500)</td>
<td></td>
<td>63</td>
</tr>
<tr>
<td>TOX3</td>
<td>rabbit pc</td>
<td>WB (1:500)</td>
<td></td>
<td>63</td>
</tr>
<tr>
<td>α-Tubulin (B-5-1-2)</td>
<td>SIGMA</td>
<td>mouse mc</td>
<td>WB (1:250), IHC</td>
<td>63</td>
</tr>
<tr>
<td>YBX-1</td>
<td>goat pc</td>
<td>WB (1:200-1:5000)</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Alexa 488-conjugated anti rabbit</td>
<td>Invitrogen Molecular Probes</td>
<td>IF (1:200)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Alexa 568-conjugated anti mouse</td>
<td>Invitrogen Molecular Probes</td>
<td>IF (1:200)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Goat IgG</td>
<td>Santa Cruz GE healthcare</td>
<td>donkey</td>
<td>WB (1:10,000)</td>
<td>-</td>
</tr>
<tr>
<td>Mouse IgG</td>
<td>Santa Cruz GE healthcare</td>
<td>donkey</td>
<td>WB (1:10,000)</td>
<td>-</td>
</tr>
<tr>
<td>Rabbit IgG</td>
<td>GE Healthcare BD</td>
<td>donkey</td>
<td>WB (1:10,000)</td>
<td>-</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>---------</td>
</tr>
<tr>
<td>IgG1-PE-Cy5 (G18-145)</td>
<td>BD Biosciences</td>
<td>mouse</td>
<td>Flow cytometry</td>
<td>-</td>
</tr>
<tr>
<td>CD10-PE-Cy7 (HI10a)</td>
<td>341112</td>
<td>mouse</td>
<td>Flow cytometry</td>
<td>-</td>
</tr>
<tr>
<td>CD24-PE (ML5)</td>
<td>555428</td>
<td>mouse</td>
<td>Flow cytometry</td>
<td>-</td>
</tr>
<tr>
<td>CD29-PE-Cy5 (MAR4)</td>
<td>559882</td>
<td>mouse</td>
<td>Flow cytometry</td>
<td>-</td>
</tr>
<tr>
<td>CD49d (9F10)</td>
<td>560972</td>
<td>mouse</td>
<td>Flow cytometry</td>
<td>-</td>
</tr>
<tr>
<td>CD49f-PE (GoH3)</td>
<td>555736</td>
<td>mouse</td>
<td>Flow cytometry</td>
<td>-</td>
</tr>
</tbody>
</table>

WB, Western blot; IHC, immunohistochemistry; IP, immunoprecipitation; IF, immunofluorescence; mc, monoclonal; pc, polyclonal.
6.3 Ethics approvals

ANNUAL REPORT PROFORMA

Annual Report Forms should be submitted to the School Ethics Committee (a subcommittee of UTREC). Please complete each section as applicable.

Name of Researcher(s): [Redacted]

Project Approval Code: MO 0922

Date Approved: 05/06/2006

Microscopy analysis of human breast cancer

The research project is still within its three year ethical approval period: YES ☑ NO ☐

I confirm that there has been no significant change to the project that may alter its ethical consideration: YES ☑ NO ☐

I confirm that, during the project, there has been no serious adverse event which has ethical implications: YES ☑ NO ☐

The research Project was completed on

Signature(s): [Signature]

cc: Supervisor (if an undergraduate or post-graduate student)

1 Note: For any significant change, a new application for ethical approval must be required.

2 Note: If there has been a serious adverse event, the School Ethics Committee must be informed as soon as possible.
ANNUAL REPORT PROFORMA

Annual Report Forms should be submitted to the School Ethics Committee (a subcommittee of UTREC),
Plus one copy to the Secretary of UTREC (timl@st-and.ac.uk)

Name of Researcher(s): [Signature(s):]

Project Approval Code: [Rb 0532] Date Approved: 15/3/07
Title: Development of analytic algorithms
Please complete each section as applicable:
Date 15/3/08

The research project is still within its three year ethical approval period: YES □ NO □

I confirm that there has been no significant change to the project that may alter its ethical consideration: YES □ NO □

I confirm that during the project, there has been no serious adverse event which has ethical implications: YES □ NO □

The research project was completed on [Date]

cc: Supervisor (if an undergraduate or post-graduate student)

Note: For any significant change, a new application for ethical approval needs to be sent. Note: If there has been a serious adverse event, the School Ethics Committee must be informed as soon as possible.

St Mary's Quad, South Street, St Andrews, Fife KY16 9AL
Tel: (03334) 462041 Fax: (01334) 463042

203
ANNUAL REPORT PROFORMA

Name of Researcher(s):

Project Approval Code: MD 1538
Date Approved: 04/12/06

The research project is still within its three year ethical approval period

I confirm that there has been no significant change to the project that may alter its ethical consideration.

I confirm that, during the project, there has been no serious adverse event which has ethical implications.

The research project was completed on

Signature(s):

cc Supervisor (if an undergraduate or post-graduate student)

1 Note: For any significant change, a new application for ethical approval needs to be made.
2 Note: If there has been a serious adverse event, the School Ethics Committee must be informed as soon as possible.

St Mary’s Quad, South Street, St Andrews, Fife KY16 9JD
Tel: (01334) 463081 Fax: (01334) 463642
Appendix

University of St Andrews

Bute Medical School Ethics Committee

6th June 2006

Professor R Iggo
Bute Medical School

Dear Richard

Microarray analysis of human breast cancer

The above application was discussed at a meeting of the ethics committee of the Bute Medical School on 5th June 2006. I am pleased to inform you that the study was approved from an ethical point of view.

In accordance with University policy, approval is granted for 3 years from the date of this meeting but this is dependent on receipt of annual progress reports on the ethical aspects of the study. Please note that, as Principal Investigator in St Andrews, you are responsible for ensuring that these reports are sent on a yearly basis to the secretary to the School ethics committee.

Please note that projects that have not commenced within two years of original approval must be re-submitted to the School ethics committee.

You must inform the School ethics committee when the research has been completed. If you are unable to complete the research within the 3 year validation period, you will be required to write to the School ethics committee to request an extension, or you will need to re-apply.

Any serious adverse events or significant change that occurs in connection with this study and/or that may alter its ethical consideration, must be reported to the School ethics committee.

Yours sincerely

C Simon Harrington MA DPhil FRCP FRCPath
Convenor, Bute Medical School Ethics Committee

cc Prof RH MacDougall, Head of School
Appendix

University of St Andrews
Bute Medical School

6th December 2006

Professor Richard Iggm
Bute Medical School

Dear Trevor,

Analysis of Mammary Tumour Models

The above application was discussed at a meeting of the Ethics committee of the Bute Medical School on 6th December 2006. I am pleased to inform you that the application was approved. In accordance with University policy, approval is granted for 3 years from the date of approval.

Please note that any modifications to the submitted protocol require ratification by the committee.

Yours sincerely

[Signature]

Professor G Humphris PhD M ClinPsychol CPsychol
Convenor Bute Medical School Ethics Committee
6.4 Mass spectrometry

6.4.1 Mass spectrometry specifications and data analysis

Traitement des échantillons

<table>
<thead>
<tr>
<th>Produits</th>
<th>Fournisseur</th>
<th>Référence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsine</td>
<td>Sigma</td>
<td>T6567</td>
</tr>
<tr>
<td>Ammonium bicarbonate</td>
<td>Sigma</td>
<td>A6141</td>
</tr>
<tr>
<td>Acetonitrile</td>
<td>Sigma R</td>
<td>Chromasolv 34881</td>
</tr>
<tr>
<td>Acide formique</td>
<td>Prolabo</td>
<td>Normapur 20 318.297</td>
</tr>
<tr>
<td>Acide acétique</td>
<td>Prolabo</td>
<td>Normapur 20 104.298</td>
</tr>
<tr>
<td>Dithiotreitol</td>
<td>Promega</td>
<td>V3151</td>
</tr>
<tr>
<td>Iodoscétemide</td>
<td>Sigma</td>
<td>I6125</td>
</tr>
<tr>
<td>Kit iTRAQ</td>
<td>Applied Biosystems</td>
<td>4352125</td>
</tr>
<tr>
<td>iChrom</td>
<td>iChrom</td>
<td>PTD/00001/64</td>
</tr>
</tbody>
</table>

- Découpage : effectué le 07/02/2011 par AM Lomenech
  - Scalpel modifications
  - Réduction/Alkylation : effectué le par choix
    - modifications
  - Digestion : effectué le 07/02/2011 par AM Lomenech
    - in gel PMQ-MS-02-07 modifications
  - Extraction : effectué le 08/02/2011 par AM Lomenech
    - in gel PMQ-MS-02-07 modifications
  - Traitement pré-MS : effectué le par choix
    - modifications
  - iTRAQ : effectué le par choix
    - modifications
  - Traitement particulier : effectué le par choix
    - modifications

- Remarque :
## Analyse MS

<table>
<thead>
<tr>
<th>Analyse MS</th>
<th>Ech.</th>
<th>Instrument (1)</th>
<th>Méthode Acquisition</th>
<th>Date Injection</th>
<th>Stockage C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>x110201_TOK_03</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_03</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_04</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_05</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_06</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_07</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_08</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_09</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_10</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_11</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_12</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_13</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_14</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_15</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
<tr>
<td>x110201_TOK_16</td>
<td>Phase Tok 3 Bande 2</td>
<td>LTQ Orbitrap</td>
<td>FT-IT_TOFS_CID.6 MT06</td>
<td>11/02/2011</td>
<td>2011.02.13</td>
</tr>
</tbody>
</table>

(1) LCQ Deca X (Thermo) ou LTQ Orbitrap (Thermo).

---

**Qualification de la MS :**

effectué le 07/02/2011 par S. Claverol

Test Michrom PMQ MS 19.03

Fichier de calibration : Michrom_Ori_110207.05

**Remarque :**

- **Méthodes d’acquisition**

  **Méthode LCQ : choix**

  - **A** : 95/05/1.1 H2O / ACN/ HCOOH v/v/v
  - **B** : 30:0:0:1 H2O / ACN / HCOOH v/v/v
  - gradient : 2-40% B en 35 minutes
  - volume injecté : 10 μL
  - précolonne : 300 um 10 x 5 mm C18 PepmapTM
  - débit précolonne : 30 μL/min
  - colonne : 75 μm ID x 15 cm C18 PepmapTM
  - débit colonne : 200 nL/min
  - needle voltage : 2 kV
  - capillary voltage : 49 V
  - psion : 1.0 en MS et 1.1 en MS
  - MS/MS strategy : MS + 2 CID MS/MS
  - min. signal required : 500
  - ps isolation window : 3 m/z units
  - normalized collision energy 35%
  - default charge state 2
  - activation Q : 0.20
  - activation time : 30
  - dynamic excitation: ON ; params 30 sec, 500; 10 pmf
  - charge state screening : ON
  - charge state rejection ON ; params

---

208
Interrogation des banques de données & validation

effectuée le 15/02/2011 par JW Dupuy

DTA generation & filtering

<table>
<thead>
<tr>
<th>MW range</th>
<th>LCQ Deca XP</th>
<th>LTQ Orbitrap</th>
<th>LTQ XL with ETD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW range</td>
<td>600-3500</td>
<td>600-4500</td>
<td>600-7500</td>
</tr>
<tr>
<td>Treshold</td>
<td>Absolue/Relative</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>precursor ion tolerance</td>
<td>1.4</td>
<td>50 ppm</td>
<td>-</td>
</tr>
<tr>
<td>group scan</td>
<td>10</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>mini. group count</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>mini. ion count</td>
<td>15</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>use charge state : auto</td>
<td>MS² level : auto</td>
<td>Activation type : CEI</td>
<td></td>
</tr>
<tr>
<td>DTA filtering : No filter</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DTA search

<table>
<thead>
<tr>
<th>Mass type</th>
<th>Monoisotopic precursor and fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzyme Units</td>
<td>Fully enzymatic (cleaves at both ends)</td>
</tr>
<tr>
<td>Miss cleav.</td>
<td>2</td>
</tr>
<tr>
<td>Peptide tolerance</td>
<td>10 ppm</td>
</tr>
<tr>
<td>Fragment ion tolerance</td>
<td>0.3 uA</td>
</tr>
<tr>
<td>Remove precursor ion from spectra</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Ions and Ion series calculated: a, b, c, x, y, z

Modifications

<table>
<thead>
<tr>
<th>PTM Per peptide</th>
<th>Symbol</th>
<th>AA</th>
<th>Mass</th>
<th>Differentiable</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>15.94/92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>C</td>
<td>57.02146</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Validation Filters

- Peptide Delta CNA: 0.1
- Peptide Xcorr vs Charge States: 1.50 (+1), 2.00 (+2), 2.50 (+3), 3.00 (+4)
- Peptide Peptide Probability: 0.001
- Peptide # Matches: 1
- Protein Number of Different Peptides: 2
- Protein Exclude this reference: keratin

Database

<table>
<thead>
<tr>
<th>Name</th>
<th>SP_Human_110215_REV_KT_xS110215</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Banque SwissProt Homo sapiens - Banque Inverse - Banque Keratines/Typsine - Séquence client</td>
</tr>
<tr>
<td>Format</td>
<td>fasta</td>
</tr>
<tr>
<td>Entries:</td>
<td>Date: 02/02/2011</td>
</tr>
<tr>
<td>Direct</td>
<td>Species</td>
</tr>
<tr>
<td>Reversed</td>
<td>Species</td>
</tr>
<tr>
<td>Total</td>
<td>Species</td>
</tr>
</tbody>
</table>

PepQuan Area/Height Parameters

Report peak Area Height

Mass tolerance 1.5 Minimum threshold 500000 Number of smoothing points 5
### 6.4.2 Mass spectrometry results including keratins

**Table 13. Mass spectrometry results including keratins**

<table>
<thead>
<tr>
<th>Protein family</th>
<th>Family rank</th>
<th>Reference</th>
<th># MS/MS Scan</th>
<th># Different Peptides</th>
<th>Coverage</th>
<th>MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1&gt; Father</td>
<td>sp</td>
<td>P04264</td>
<td>K2C1_HUMAN Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=6 TAP-TDI full length (N=C: CBP-spacer-SBP-linker-TDX3)</td>
<td>523</td>
<td>31</td>
<td>45.65217</td>
</tr>
<tr>
<td>&lt;2&gt; Father</td>
<td>sp</td>
<td>P35527</td>
<td>K2C5_HUMAN Keratin, type I cytoskeletal 9 - Homo sapiens sp</td>
<td>P35527</td>
<td>K2C5_HUMAN Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3 P13645</td>
<td>K1C10_HUMAN Keratin, type I cytoskeletal 10 OS=Homo sapiens sp</td>
</tr>
<tr>
<td>&lt;5&gt; Father</td>
<td>sp</td>
<td>P62979</td>
<td>RS27A_HUMAN Ubiquitin Q5XQN5</td>
<td>K2C5_BOVIN Keratin, type II cytoskeletal 15 OS=Homo sapiens GN=KRT6 PE=1 SV=6 sp</td>
<td>P35980</td>
<td>K22E_HUMAN Keratin, type II cytoskeletal 2 epidermal - Homo sapiens</td>
</tr>
<tr>
<td>&lt;6&gt; Independant</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 14 OS=Homo sapiens sp</td>
<td>P02533</td>
<td>K1C14_HUMAN Keratin, type I cytoskeletal 14 OS=Homo sapiens sp</td>
<td>K1C14_HUMAN Keratin, type I cytoskeletal 14 OS=Homo sapiens</td>
</tr>
<tr>
<td>&lt;11&gt; Twin</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 14 OS=Homo sapiens sp</td>
<td>P02533</td>
<td>K1C14_HUMAN Keratin, type I cytoskeletal 14 OS=Homo sapiens sp</td>
<td>K1C14_HUMAN Keratin, type I cytoskeletal 14 OS=Homo sapiens</td>
</tr>
<tr>
<td>&lt;12&gt; Father</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 16 OS=Homo sapiens sp</td>
<td>K1C14_HUMAN Keratin, type I cytoskeletal 16 OS=Homo sapiens sp</td>
<td>K1C14_HUMAN Keratin, type I cytoskeletal 16 OS=Homo sapiens</td>
<td>126</td>
</tr>
<tr>
<td>&lt;13&gt; Father</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 16 OS=Homo sapiens sp</td>
<td>K1C14_HUMAN Keratin, type I cytoskeletal 16 OS=Homo sapiens sp</td>
<td>K1C14_HUMAN Keratin, type I cytoskeletal 16 OS=Homo sapiens</td>
<td>109</td>
</tr>
<tr>
<td>&lt;14&gt; Father</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 16 OS=Homo sapiens sp</td>
<td>K1C14_HUMAN Keratin, type I cytoskeletal 16 OS=Homo sapiens sp</td>
<td>K1C14_HUMAN Keratin, type I cytoskeletal 16 OS=Homo sapiens</td>
<td>105</td>
</tr>
<tr>
<td>&lt;15&gt; Independant</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 14 - Mus musculus P13647</td>
<td>K2C5_HUMAN Keratin, type II cytoskeletal 5 - Homo sapiens sp</td>
<td>P13647</td>
<td>K2C5_HUMAN Keratin, type II cytoskeletal 5 OS=Homo sapiens GN=KRT5 PE=1 SV=3 sp</td>
</tr>
<tr>
<td>&lt;19&gt; Father</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 5 - Mus musculus P00761</td>
<td>TRYP_Pig Trypsin precursor - Sus scrofa sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 5 - Mus musculus</td>
</tr>
<tr>
<td>&lt;22&gt; Independant</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 5 - Mus musculus P00761</td>
<td>TRYP_Pig Trypsin precursor - Sus scrofa sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 5 - Mus musculus</td>
</tr>
<tr>
<td>&lt;23&gt; Independant</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 5 - Mus musculus P00761</td>
<td>TRYP_Pig Trypsin precursor - Sus scrofa sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type I cytoskeletal 5 - Mus musculus</td>
</tr>
<tr>
<td>&lt;24&gt; Twin</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type II cytoskeletal 5 - Bos taurus sp</td>
<td>P62979</td>
<td>RPS27A_HUMAN Ubiquitin-4OS ribosomal protein S27a OS=Homo sapiens GN=RPS27A PE=1 SV=2 sp</td>
<td>P62979</td>
</tr>
<tr>
<td>&lt;25&gt; Twin</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type II cytoskeletal 5 - Bos taurus sp</td>
<td>P62979</td>
<td>RPS27A_HUMAN Ubiquitin-4OS ribosomal protein S27a OS=Homo sapiens GN=RPS27A PE=1 SV=2 sp</td>
<td>P62979</td>
</tr>
<tr>
<td>&lt;26&gt; Twin</td>
<td>sp</td>
<td>Q61781</td>
<td>K1C14_MOUSE Keratin, type II cytoskeletal 5 - Bos taurus sp</td>
<td>P62979</td>
<td>RPS27A_HUMAN Ubiquitin-4OS ribosomal protein S27a OS=Homo sapiens GN=RPS27A PE=1 SV=2 sp</td>
<td>P62979</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>q-value</td>
<td>Fragment</td>
<td>Protein</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>---------</td>
<td>----------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>P160797</td>
<td>hIL-13</td>
<td>0.001</td>
<td>47.2144</td>
<td>IL-13</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>P160800</td>
<td>hIL-14</td>
<td>0.002</td>
<td>39.1426</td>
<td>IL-14</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>P160801</td>
<td>hIL-15</td>
<td>0.003</td>
<td>31.2345</td>
<td>IL-15</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td>P160802</td>
<td>hIL-16</td>
<td>0.004</td>
<td>23.3456</td>
<td>IL-16</td>
<td>0.0004</td>
<td></td>
</tr>
<tr>
<td>P160803</td>
<td>hIL-17</td>
<td>0.005</td>
<td>15.4567</td>
<td>IL-17</td>
<td>0.0005</td>
<td></td>
</tr>
<tr>
<td>P160804</td>
<td>hIL-18</td>
<td>0.006</td>
<td>8.5678</td>
<td>IL-18</td>
<td>0.0006</td>
<td></td>
</tr>
<tr>
<td>P160805</td>
<td>hIL-19</td>
<td>0.007</td>
<td>6.7890</td>
<td>IL-19</td>
<td>0.0007</td>
<td></td>
</tr>
<tr>
<td>P160806</td>
<td>hIL-20</td>
<td>0.008</td>
<td>4.9012</td>
<td>IL-20</td>
<td>0.0008</td>
<td></td>
</tr>
<tr>
<td>P160807</td>
<td>hIL-21</td>
<td>0.009</td>
<td>3.1234</td>
<td>IL-21</td>
<td>0.0009</td>
<td></td>
</tr>
</tbody>
</table>

**Table 13. continued**

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>q-value</th>
<th>Fragment</th>
<th>Protein</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>P160808</td>
<td>hIL-22</td>
<td>0.010</td>
<td>2.3456</td>
<td>IL-22</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

**Appendix**

**Table 13. continued**

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>q-value</th>
<th>Fragment</th>
<th>Protein</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>P160809</td>
<td>hIL-23</td>
<td>0.011</td>
<td>1.5678</td>
<td>IL-23</td>
<td>0.0011</td>
</tr>
<tr>
<td>P160810</td>
<td>hIL-24</td>
<td>0.012</td>
<td>0.7890</td>
<td>IL-24</td>
<td>0.0012</td>
</tr>
<tr>
<td>P160811</td>
<td>hIL-25</td>
<td>0.013</td>
<td>0.5678</td>
<td>IL-25</td>
<td>0.0013</td>
</tr>
<tr>
<td>P160812</td>
<td>hIL-26</td>
<td>0.014</td>
<td>0.3456</td>
<td>IL-26</td>
<td>0.0014</td>
</tr>
<tr>
<td>P160813</td>
<td>hIL-27</td>
<td>0.015</td>
<td>0.2345</td>
<td>IL-27</td>
<td>0.0015</td>
</tr>
<tr>
<td>P160814</td>
<td>hIL-28</td>
<td>0.016</td>
<td>0.1234</td>
<td>IL-28</td>
<td>0.0016</td>
</tr>
<tr>
<td>P160815</td>
<td>hIL-29</td>
<td>0.017</td>
<td>0.0345</td>
<td>IL-29</td>
<td>0.0017</td>
</tr>
<tr>
<td>P160816</td>
<td>hIL-30</td>
<td>0.018</td>
<td>0.0123</td>
<td>IL-30</td>
<td>0.0018</td>
</tr>
</tbody>
</table>
Table 13. continued

<table>
<thead>
<tr>
<th>#</th>
<th>Protein</th>
<th>Parent or Subunit</th>
<th>Accession</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>27=</td>
<td>9=</td>
<td>16.2914</td>
<td>57249.88</td>
<td></td>
</tr>
<tr>
<td>27=</td>
<td>9=</td>
<td>16.2914</td>
<td>57249.88</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>9</td>
<td>42.5925</td>
<td>35074.2</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>5.70576</td>
<td>35057.59</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>10</td>
<td>27.7027</td>
<td>49953.94</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>7</td>
<td>19.1111</td>
<td>50400.28</td>
<td></td>
</tr>
<tr>
<td>20=</td>
<td>19=</td>
<td>8.700481</td>
<td>29564.85</td>
<td></td>
</tr>
<tr>
<td>20=</td>
<td>19=</td>
<td>9.00932</td>
<td>31681.73</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>19=</td>
<td>8.106243</td>
<td>31583.8</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>19=</td>
<td>8.245614</td>
<td>31254.04</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>19=</td>
<td>8.339248</td>
<td>30851.42</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>9</td>
<td>26.94611</td>
<td>36615.16</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>5.806452</td>
<td>35148.81</td>
<td></td>
</tr>
<tr>
<td>18=</td>
<td>7=</td>
<td>19.04762</td>
<td>50109.18</td>
<td></td>
</tr>
<tr>
<td>18=</td>
<td>7=</td>
<td>19.04762</td>
<td>50153.22</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>9</td>
<td>16.2914</td>
<td>57249.88</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>5</td>
<td>4.769475</td>
<td>64464.63</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>5</td>
<td>5.89172</td>
<td>64737.7</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>3.17965</td>
<td>64301.73</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>16</td>
<td>4.412398</td>
<td>53207.18</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>13</td>
<td>18.30303</td>
<td>90528.01</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>6.636156</td>
<td>47724.3</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>9</td>
<td>44.85997</td>
<td>66711.44</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>25.625</td>
<td>34175.24</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>22.04301</td>
<td>38723.01</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>30.71161</td>
<td>29368.24</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>9.392265</td>
<td>58999.61</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>25.87904</td>
<td>73070.13</td>
<td></td>
</tr>
</tbody>
</table>

Shown are the top ranking peptides according to the number of scans.
### 6.5 Gene set enrichment analysis of the TOX3 microarray

Table 14. Top ranked genes in TOX3 arrays

<table>
<thead>
<tr>
<th>Gene</th>
<th>Enrichment score</th>
<th>Gene</th>
<th>Enrichment score</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP2</td>
<td>2.1259015</td>
<td>BCYRN1</td>
<td>0.39984468</td>
</tr>
<tr>
<td>HLA-DRB1</td>
<td>1.1881273</td>
<td>FOXD2</td>
<td>0.39857835</td>
</tr>
<tr>
<td>KRT86</td>
<td>1.1044508</td>
<td>LGALS8</td>
<td>0.39836216</td>
</tr>
<tr>
<td>TGFBI</td>
<td>1.0102248</td>
<td>TRIP13</td>
<td>0.3976165</td>
</tr>
<tr>
<td>HLA-DRB4</td>
<td>0.914595</td>
<td>TFF2</td>
<td>0.3975265</td>
</tr>
<tr>
<td>ACOX2</td>
<td>0.90986234</td>
<td>INHBB</td>
<td>0.3942466</td>
</tr>
<tr>
<td>SLC16A6</td>
<td>0.87592816</td>
<td>LOC399942</td>
<td>0.39258313</td>
</tr>
<tr>
<td>CXL14</td>
<td>0.85513693</td>
<td>CDCA5</td>
<td>0.3919771</td>
</tr>
<tr>
<td>HLA-DRB3</td>
<td>0.8402379</td>
<td>SLC29A1</td>
<td>0.3892715</td>
</tr>
<tr>
<td>HS.750995</td>
<td>0.8361645</td>
<td>ITGB5</td>
<td>0.3891685</td>
</tr>
<tr>
<td>HLA-DRB6</td>
<td>0.78940094</td>
<td>ITPR1PL2</td>
<td>0.3896616</td>
</tr>
<tr>
<td>HLAF5</td>
<td>0.7835287</td>
<td>CLDN23</td>
<td>0.38855027</td>
</tr>
<tr>
<td>LOC145837</td>
<td>0.7684144</td>
<td>HRASLS3</td>
<td>0.38900054</td>
</tr>
<tr>
<td>ACACB</td>
<td>0.6672306</td>
<td>HNRNPA1</td>
<td>0.38854802</td>
</tr>
<tr>
<td>MT2A</td>
<td>0.63128906</td>
<td>PRR11</td>
<td>0.38745415</td>
</tr>
<tr>
<td>UCK2</td>
<td>0.62720376</td>
<td>LOC642031</td>
<td>0.3831716</td>
</tr>
<tr>
<td>ENDOGL1</td>
<td>0.62679964</td>
<td>ECHS1</td>
<td>0.37764066</td>
</tr>
<tr>
<td>MT1A</td>
<td>0.6104609</td>
<td>FJX1</td>
<td>0.37719676</td>
</tr>
<tr>
<td>IFT122</td>
<td>0.6014796</td>
<td>SLC25A1</td>
<td>0.3763162</td>
</tr>
<tr>
<td>CYP24A1</td>
<td>0.6012127</td>
<td>SYT13</td>
<td>0.37492865</td>
</tr>
<tr>
<td>HBA2</td>
<td>0.5927987</td>
<td>RPP40</td>
<td>0.37312594</td>
</tr>
<tr>
<td>APOD</td>
<td>0.59222525</td>
<td>SRM</td>
<td>0.37118223</td>
</tr>
<tr>
<td>DUT</td>
<td>0.58851224</td>
<td>C120RF24</td>
<td>0.36954233</td>
</tr>
<tr>
<td>FGFBP2</td>
<td>0.5743703</td>
<td>SRPRB</td>
<td>0.36862534</td>
</tr>
<tr>
<td>C2DORF100</td>
<td>0.5730592</td>
<td>TMEM158</td>
<td>0.36776766</td>
</tr>
<tr>
<td>CRTC1</td>
<td>0.5602652</td>
<td>PCSK6</td>
<td>0.36752793</td>
</tr>
<tr>
<td>CXC4R3</td>
<td>0.5372352</td>
<td>EXOSC5</td>
<td>0.3642964</td>
</tr>
<tr>
<td>KRTB1</td>
<td>0.529781</td>
<td>LOC652903</td>
<td>0.36298683</td>
</tr>
<tr>
<td>DNYPEP</td>
<td>0.5239767</td>
<td>EFDH2</td>
<td>0.36079088</td>
</tr>
<tr>
<td>BTG2</td>
<td>0.5127981</td>
<td>CD320</td>
<td>0.3599116</td>
</tr>
<tr>
<td>HES1</td>
<td>0.5082694</td>
<td>MSMB</td>
<td>0.35867983</td>
</tr>
<tr>
<td>TTT12</td>
<td>0.50128824</td>
<td>ECHDC3</td>
<td>0.3583621</td>
</tr>
<tr>
<td>MUC1L</td>
<td>0.49363706</td>
<td>SCN1B</td>
<td>0.35690814</td>
</tr>
<tr>
<td>C14CD86</td>
<td>0.49300033</td>
<td>MTHFD1L</td>
<td>0.35689217</td>
</tr>
<tr>
<td>SERPINA3</td>
<td>0.46371156</td>
<td>MCM2</td>
<td>0.35475674</td>
</tr>
<tr>
<td>CDT1</td>
<td>0.4635114</td>
<td>MT1F</td>
<td>0.3534334</td>
</tr>
<tr>
<td>BMP7</td>
<td>0.44970667</td>
<td>GPRC5A</td>
<td>0.3531575</td>
</tr>
<tr>
<td>MT1X</td>
<td>0.44575307</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BGN</td>
<td>0.43882295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sep.05</td>
<td>0.43523532</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WWC3</td>
<td>0.43057227</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITX1</td>
<td>0.43279195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN11A</td>
<td>0.43040892</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM4SF1</td>
<td>0.42984325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC441019</td>
<td>0.42983177</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAL</td>
<td>0.42867112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LINGO1</td>
<td>0.42225155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDC45L</td>
<td>0.42196092</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPM1G</td>
<td>0.42102593</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYEOP</td>
<td>0.41971403</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNORD13</td>
<td>0.4197067</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGFBRAP1</td>
<td>0.41917968</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRPM5</td>
<td>0.41780734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASNS</td>
<td>0.41679037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYC</td>
<td>0.41395298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H19</td>
<td>0.41376397</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCNF</td>
<td>0.40846977</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOP1</td>
<td>0.4072227</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC648638</td>
<td>0.40501574</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DRA</td>
<td>0.40302932</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NKD2</td>
<td>0.40237397</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 14. continued

<table>
<thead>
<tr>
<th>Gene</th>
<th>Enrichment score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMAIP1</td>
<td>-0.3431231</td>
</tr>
<tr>
<td>DHR57</td>
<td>-0.34358507</td>
</tr>
<tr>
<td>TEMEM219</td>
<td>-0.3451259</td>
</tr>
<tr>
<td>CD68</td>
<td>-0.34623927</td>
</tr>
<tr>
<td>MCL1</td>
<td>-0.34807625</td>
</tr>
<tr>
<td>ALPP</td>
<td>-0.34889764</td>
</tr>
<tr>
<td>ADSS1L1</td>
<td>-0.34993732</td>
</tr>
<tr>
<td>GBP2</td>
<td>-0.35209325</td>
</tr>
<tr>
<td>KIA1598</td>
<td>-0.35334307</td>
</tr>
<tr>
<td>HS.72200</td>
<td>-0.3539912</td>
</tr>
<tr>
<td>NFKB1Z</td>
<td>-0.35542876</td>
</tr>
<tr>
<td>APH1B</td>
<td>-0.35582617</td>
</tr>
<tr>
<td>ARL6IP5</td>
<td>-0.35584757</td>
</tr>
<tr>
<td>PKB</td>
<td>-0.35925603</td>
</tr>
<tr>
<td>HS.570308</td>
<td>-0.36087173</td>
</tr>
<tr>
<td>RAB31</td>
<td>-0.36808662</td>
</tr>
<tr>
<td>C9ORF61</td>
<td>-0.36810032</td>
</tr>
<tr>
<td>TNFRSF12A</td>
<td>-0.36817186</td>
</tr>
<tr>
<td>ISG20</td>
<td>-0.36875505</td>
</tr>
<tr>
<td>LMBR1</td>
<td>-0.36910032</td>
</tr>
<tr>
<td>DYNTL3</td>
<td>-0.3691502</td>
</tr>
<tr>
<td>PLK2</td>
<td>-0.36918692</td>
</tr>
<tr>
<td>MGST3</td>
<td>-0.36930586</td>
</tr>
<tr>
<td>HS.25318</td>
<td>-0.369337926</td>
</tr>
<tr>
<td>NMD3</td>
<td>-0.36937685</td>
</tr>
<tr>
<td>LOC643431</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>ISG20</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>RDX</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>CLIC3</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>HIST1H3D</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>LOC653610</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>POLR3GL</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>SLC39A8</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>PAM</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>TM7SF2</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>CDKN1A</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>DUSP4</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>SLC6A14</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>S100A14</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>NCAM2</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>GPR126</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>ENTPD8</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>CPE</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>RNFT1</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>CD44</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>RN7SK</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>SSFA2</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>LOC493869</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>S100A7</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>AKR1C3</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>ERP27</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>ELP3</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>HS.571151</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>INPP4B</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>TP53INP1</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>FAM83B</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>S100A6</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>FAM113B</td>
<td>-0.369400775</td>
</tr>
<tr>
<td>BN1PL</td>
<td>-0.369400775</td>
</tr>
</tbody>
</table>

The 100 top and bottom ranking genes in the TOX3 arrays were selected based on their GSEA enrichment score.