Université de Montréal

Engrailed genes in mammary development and tumorigenesis

par

Nicole Martin

Programme de biologie moléculaire, Université de Montréal
Faculté des études supérieures

Thèse présentée à la Faculté des études supérieures
En vue de l'obtention du grade de doctorat
En biologie moléculaire

May, 2006

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

***Engrailed* genes in mammary development and tumorigenesis**

présentée par

Nicole Martin

a été évaluée par un jury composé des personnes suivantes :

Dr. Trang Hoang
Présidente-rapporteuse

Dr. Guy Sauvageau
Directeur de recherche

Dr. Sylvie Mader
Membre du Jury

Dr. Louis Gaboury
Membre du Jury

Dr. William Muller
Examinateur externe
Université de McGill

Thèse acceptée le:.....

RESUME EN FRANÇAIS ET MOTS CLES FRANÇAIS

Les gènes homéotiques *Engrailed-1 (En-1)* et *Engrailed-2 (En-2)* sont d'importants régulateurs du développement du système nerveux central chez la souris. Leur implication dans le développement de la glande mammaire et leur contribution à la transformation de ce tissu demeurent inconnu. Le travail présenté dans cette thèse explore une approche de gène candidat afin de répondre aux questions suivantes: (i) Est-ce que ces gènes sont impliqués dans le développement du tissu mammaire? (ii) Est-ce que ces gènes participent à la transformation néoplasique du sein chez l'humain? Mes résultats démontrent que bien que l'expression d'*Engrailed-2* n'est pas détectable dans le tissu mammaire murin au cours du développement fétal et chez la souris adulte, le gène paralogue *En-1* lui est fortement exprimé dans l'épithélium mammaire de souris pré-pubaires. Mes travaux démontrent aussi que les souris chez qui le gène *En-1* est inactivé par recombinaison homologue présente une glande mammaire anormale caractérisée par une rareté d'arborisation des canaux mammaires. Des expériences de transplantations de glandes mammaires isolées de souris *En-1* mutantes confirment que ces observations sont probablement le résultat d'un défaut intrinsèque à l'arbre mammaire mutant. Ces études démontrent donc pour la première fois l'expression spatio-temporel d'*Engrailed-1* dans la glande mammaire de souris et suggèrent un rôle important pour ce gène dans le développement mammaire. Mes travaux ont aussi démontrés que *EN2* (mais pas *EN1*) est exprimé de façon ectopique dans environ 8% des cancer mammaires humains et chez près de la moitié des lignées cellulaires malignes. La sur-expression d'*EN2* induit la transformation mammaire chez deux lignées non-transformées *in vitro* et la genèse d'adénocarcinomes *in vivo*. Des études d'interférence à l'ARN (RNAi) ont démontré qu'*EN2* est requis pour la prolifération de cellules transformées humaines. De plus, des analyses par micro-puces effectuées chez 2 lignées cellulaires chez qui les niveaux d'EN2 furent expérimentalement modifiés ont permis d'identifions des gènes cibles potentiels de cette protéine. Ces résultats suggèrent qu'EN2 est impliqué dans la transformation du tissu mammaire humain et d'EN1 contribue au développement de ce tissu.

Mots clés: *Engrailed*, homéodomaine, glande mammaire, cancer du sein, oncogène, morphogénèse ductale, transplantation, puberté, RNAi, micro-puce

RESUME EN ANGLAIS ET MOTS CLES ANGLAIS

The homeobox genes *Engrailed-1 (En-1)* and *Engrailed-2 (En-2)* occupy a prominent position in the developmental regulatory hierarchy and have been studied extensively in embryonic development, yet have received little attention with respect to mammary gland organogenesis and cancer. The studies presented in this thesis are the result of a candidate gene approach where the expression and potential role of these homeodomain-containing proteins were investigated in the developing mammary gland and in breast tumors. While *En-2* was never detected at any developmental timepoint in normal mouse or human breast tissue, we have defined the developmentally regulated expression of *En-1* in the mammary epithelium of the prepubertal and early pubertal mouse mammary gland. Moreover, using *En-1* mutant mice, we provide evidence that loss of En-1 function in the female mammary gland results in severely impaired ductal growth. Pubertal *En-1* null mammary glands revealed a primitive ductal rudiment devoid of terminal end buds (TEBs), reminiscent of a prepubertal mammary gland, while a fully developed ductal system was seen in *En-1* heterozygous and wildtype siblings. *En-1* null mammary epithelium transplanted into surgically cleared fat pads of syngeneic hosts displayed limited ductal outgrowth and a decrease in side branching. These studies demonstrate a unique spatio-temporal pattern of *En-1* expression in mammary tissue and suggest a potential role for *En-1* in the initial growth and morphogenesis of the epithelial ductal system during the onset of puberty. We also show that *EN2* (but not *EN1*) is ectopically expressed in a subgroup of human breast tumors and in a large proportion of breast cancer cell lines and that its ectopic expression readily transforms mammary epithelial cells *in vitro* and promotes adenocarcinoma formation *in vivo*. RNA interference studies show that *EN2* expression is required for the maintenance of the transformed phenotype of human breast tumor cells. Moreover, microarray analysis of both gain-of-function and loss-of-function of *EN2* in two breast cancer cell lines provided initial insight into putative EN2 responsive targets. These studies reveal that *En* genes have not only acquired a role in the postnatal development of the mouse mammary gland, but in addition, they have evolved to contribute to breast tumorigenesis.

Mots clés: *Engrailed*, homeobox, mammary gland, breast cancer, oncogene, ductal morphogenesis, transplantation, puberty, RNAi, microarray

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES ET ABRÉVIATIONS

Abréviation	Signification
17-AAG	17-(Allylamino)-17-demethoxygeldanamycin; $C_{31}H_{43}N_3O_8$
2ki	*Engrailed-2* knock-in
a.a.	Amino acid
Å	angstrom
A	Adenine; purine base $C_5H_5N_5$
ABC	Avidin: Biotinylated enzyme Complex
Ade	Adenine
Ag	Antigen
AP	Alveolar Precursor
ATP	Adenosine triphosphate
bp	base pair
β-gal	*Beta*-galactosidase
β-ME	*Beta*-mercaptoethanol
BSA	Bovine Serum Albumin
cDNA	Complementary DNA
C	Cytosine; pyrimidine base $C_4H_5N_3O$
CDS	Coding sequence
ChIP	Chromatin Immunoprecipitation
CIHR	Canadian Institute of Health Research
cDNA	Complementary DNA
cRNA	Complementary RNA
DIP	Dexamethasone, Insulin and Prolactin
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dki	*Drosophila engrailed* knock-in
dpc	days post coitum
DP	Ductal Precursor
ECM	Extracellular Matrix
EGF	Epidermal Growth Factor

EGFP	Enhanced Green Fluorescent Protein
EH	Engrailed Homology region
en	*Drosophila engrailed* gene
En-1	Murine *Engrailed-1* gene
En-2	Murine *Engrailed-2* gene
EN1	Human *Engrailed-1* gene
EN2	Human *Engrailed-2* gene
EPC	Epithelial Precursor Cell
ESA	Epithelial-specific antigen
EST	Expressed Sequence Tag
FACS	Fluorescence Activated Cell Sorting
FCS	Fetal Calf Serum
FITC	Fluorescein isothiocyanate
G	Guanine; purine base $C_5H_5N_5O$
GFP	Green Fluorescent Protein
GR	Glucocorticoid receptor
HAM	Ham's F12 Medium
hd	homeodomain deletion
HOM-C	*Drosophila* Homeotic gene complex
HOX	vertebrate HOXA-D clusters containing Homeobox genes
HRP	Horseradish Peroxidase
HSC	Hemopoietic Stem Cell
IgG	Immunoglobulin G
IRCM	Institut de Recherches Cliniques de Montréal
IVT	*In vitro* transcription
K6	Cytokeratin 6
K18	Keratin 18
kDa	kilodalton
L15	Leibovitz L15 tissue culture medium
lacZ	*E. coli lacZ* gene; coding for the enzyme β-galactosidase
LCM	Laser Capture Microdissection
lki	*lacZ* knock-in
LN	Lymph Node

LTC-IC	Long-Term Culture Initiating Cell
LTR	Long Terminal Repeat
MEF	Mouse Embryonic Fibroblast
MHC	Major Histocompatibility Complex
MMTV	Mouse Mammary Tumor Virus
mRNA	messenger RNA
MSC	Mammary Stem Cell
MSCV	Murine Stem Cell Virus
MUC1	Mucin 1, epithelial membrane antigen
N	Normal Tissue
Neo^r	Neomycin resistance gene
N.D.	Not Determined
NLS	Nuclear Localization Signal
No R.T.	No Reverse Transcriptase/No Reverse Transcription
NP40	Nonidet P-40 detergent
NT	Not Transfected/Not Transduced
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate-Buffered Salt
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PGK	Phosphoglycerate kinase
PI	Propidium Iodine
PTP-1D	Protein-tyrosine phosphatase 1D
$Puro^r$	Puromycin resistance gene
RMA	Robust Multi-array Average
RNA	Ribonucleic acid
RNAi	RNA Interference
rRNA	ribosomal RNA
RPMI	RPMI-1640 tissue culture medium
PR	Progesterone Receptor
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulfate
siCTRL	Small interfering control scrambled RNA

siRNA	Small interfering RNA
SPF	Specific Pathogen-Free
T	Thymine; pyrimidine base $C_5H_6N_2O_2$
TCF	T-Cell Factor
TdT	Terminal deoxynucleotidyl transferase
TEB	Terminal end bud
U	Uracil; pyrimidine base $C_4H_4N_2O_2$
VSV	Vesicular Stomatitis Virus
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

LA DÉDICACE

A ma famille, mes amis et mon fiancé,

REMERCIEMENTS

I would like to thank my thesis director, Dr Guy Sauvageau, for his continual encouragement and guidance. He is an exemplary Clinician Scientist who unwaveringly dedicates so much of his time and energy to his work and the well being of the lab in general, inspiring all those working under him to aspire to the same level of commitment. His enthusiasm for science is truly contagious and his ability to always view experimental challenges and setbacks as invaluable learning experiences is instrumental in many of the projects in the lab persevering and eventually flourishing as they do.

I would also like to thank all the current and past members of the Sauvageau lab for their guidance, training, scientific discussions and equally importantly, their camaraderie. I have benefited greatly from the opportunity to work alongside such inspiring and uplifting colleaugues. I would also like to thank the Canadian Institute of Health (CIHR) and l'Institut de Recherches Cliniques de Montréal (IRCM) for their financial support during my studies.

Last, but certainly not least, I would have never accomplished such an ambitious endeavor and benefitted from such an invaluable educational experience without the momentous encouragement and support from my family, friends and fiancé.

CHAPITRE 1

INTRODUCTION

Investigating a Potential Role for *Engrailed* Genes in Mammary Gland Development and Tumorigenesis

The present Ph.D. thesis, consisting of five chapters, describes the expression, function and initial mechanistic insights of action of *Engrailed-1 (En-1)* and *Engrailed-2 (En-2)* in normal mammary gland development and mammary tumorigenesis.

Chapter 1 is a literature review which shall both encompass the most relevant studies from other groups to gain a further appreciation of this field and help identify the underlying motivation behind the research contained herein and is subdivided into six sections. The first section briefly describes Homeobox genes in development, leading to the second section, which specifically describes the *Engrailed* homeobox gene family. The third section outlines the different stages of mouse mammary gland development and imparts how the mouse is an invaluable model with which to investigate genes involved in mammary gland development. The fourth section highlights the role of exemplary homeobox genes in the developing mammary gland while the fifth section summarizes the role of certain homeobox genes in cancer, with an emphasis in their potential involvement in breast cancer. The sixth section introduces the area of human breast cancer in general. The Introduction was written by Nicole Martin under the supervision of Dr. Guy Sauvageau.

1.1 Homeobox genes in embryonic development

1.1.1 Homeobox genes are key developmental regulators

Homeobox genes comprise a large family of transcriptional regulatory proteins that are involved in a wide range of essential biological processes from embryonic development to terminal differentiation. They are defined by the presence of a 180-bp DNA sequence motif designated the homeobox. The homeobox was initially characterized as a sequence motif that was shared among *Drosophila* Homeotic genes (the HOM-C complex). The homeobox is now known to be evolutionarily conserved among many genes with over 1,000 homeobox genes having been identified in several species, ranging from hydra to humans as well as fungi and plants [1]. A vast extent of what is presently known about homeobox genes and their corresponding functions is the culmination of decades of research that started with dissecting *Drosophila* development and genetics. The HOM-C genes in *Drosophila* were originally identified as part of a hierarchy of genes that control embryonic development and play key roles in the determination and maintenance of cell fate and cell identity. Mutations in these *Drosophila* HOM-C or homeotic genes result in the conversion of one body part or segment to the likeness or identity of another, and led to the coining of the term 'homeotic transformation'. Other classes of genes in this hierarchy, such as the gap, pair-ruled and segment polarity genes, which in turn play essential roles in the determination and maintenance of cell fate and pattern formation in the developing *Drosophila* embryo, also contain members that are homeobox genes.

In mammals, homeobox genes reign over the specification of the overall body plan and are known to play crucial roles in a variety of developmental processes encompassing central nervous system and skeletal development, limb and digit specification, and organogenesis. The HOM-C genes are in many respects still considered to be the prototype homeobox genes and while they represent only a subset of all known or predicted homeobox genes, their counterparts in mammals, the HOX genes, are among the most extensively studied family among vertebrate homeobox genes. Over evolutionary time, the number of homeobox genes has increased and their functions

have been reengineered to meet the demands of increasingly diverse developmental processes. To date, it is estimated that the human genome contains at least 200 known or predicted homeobox genes, of which only 39 are members of the HOX family [2]. In mammals, mutations in homeobox genes can cause dramatic developmental defects including loss of specific structures as well as classical homeotic transformations. Some homeobox genes appear to have cell autonomous functions in differentiation and cell cycle control while others appear to have non-cell autonomous functions such as pattern formation and mediation of reciprocal tissue interactions [3].

1.1.2 The homeodomain encodes a helix-turn-helix DNA-binding domain

The highly conserved 180-bp homeobox DNA sequence encodes the homeodomain, a 60 amino acid long DNA-binding domain that folds into three α-helices and a flexible N-terminal arm. Helices 2 and 3 form a helix-turn-helix motif that is the hallmark of all homeodomain-containing proteins. The third α helix, also known as the recognition helix, contacts specific bases in the major groove while the N-terminal extension contacts specific bases in the minor groove. The original 2.8 Å resolution structure of the engrailed homeodomain-DNA complex was the first crystal structure to reveal how this motif recognized DNA [4]. Although there is considerable variation in the primary sequence of the homeobox, the consensus amino acids, which are invariant among all homeodomains, maintain the overall fold and DNA-docking arrangement in the three-dimensional structure of the homeodomain (Fig. 1.1).

Evolutionary relationships and family classifications of homeobox genes are based on the level of similarity among their respective homeodomains and subsequently, by comparative analyses of amino acid sequences both amino-terminal and carboxyl-terminal to the homeodomain, which vary considerably from protein to protein. These families vary in size from the relatively large ones, such as the HOX family, that comprises 39 members, to the small families, such as the Engrailed family, which only has two members. The a.a. sequence diversity among homeodomain proteins is believed

to contribute to their distinct functional properties, by generating distinct DNA-binding specificities, promoting unique protein-protein interactions, and other mechanisms.

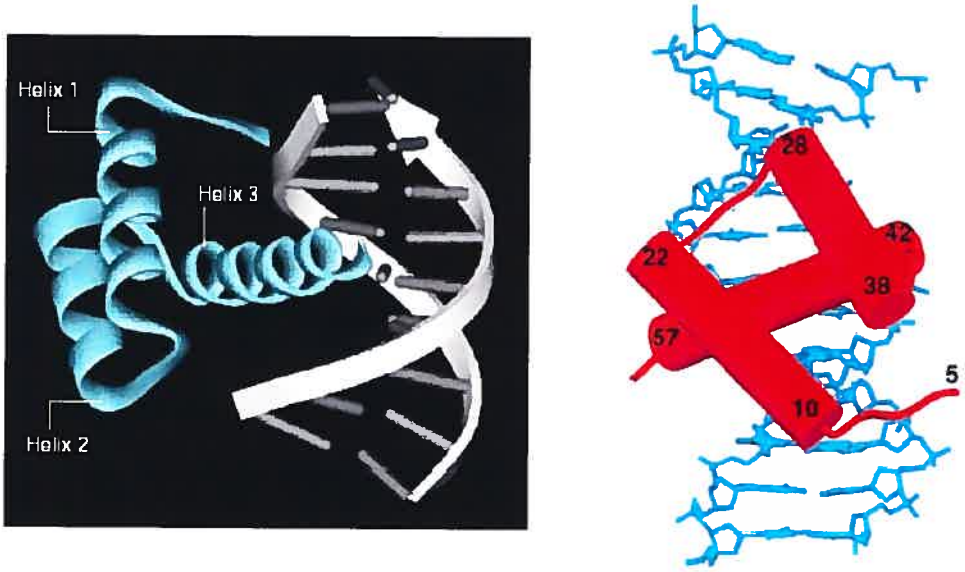

Fig. 1.1 The highly conserved three-dimensional helix-turn-helix structure encoded by the homeodomain is the hallmark of all homeodomain-containing proteins. The left image depicts the *Drosophila* engrailed homeodomain-DNA complex where the 3 α-helices and the extended N-terminal arm of the engrailed homeodomain are shown in blue. The right image shows the complementary 2.2 Å resolution structure of the *Drosophila* engrailed homeodomain bound to its optimal DNA site. DNA is depicted in blue and the protein backbone is depicted in red with α-helices represented by cylinders. Residue 5 is the first amino acid that could be reliably modeled from the crystal structure while the other numbers indicate helix termini. Adapted from Fraenkel E., Rould M.A., Chambers K.A. and Pabo C.O. (1998) *J Mol Biol.* 284:351-361.

1.1.3 Homeoproteins are transcriptional regulators

In accordance with their role as transcriptional regulators, homeobox genes encode transcription factors, which are predominantly localized in the nucleus where they have been shown to function as activators, repressors or both. These homeoproteins are thought to instruct the cell as to which genetic program they should further implement,

depending on which step these cells have reached in the course of their development, or in response to developmental cues. Although it is widely accepted that the binding promiscuity of homeoproteins *in vitro* vs. their highly selective functions *in vivo*, reflects their requirement for cofactors, relatively few examples exist in which bona fide target genes have been identified and are regulated by specific homeobox genes *in vivo*. It is now thought that the functional specificities of homeoproteins are dictated by several tiers of regulation, including post-transcriptional controls, nuclear-cytoplasmic transport and protein-protein interactions. For example, PBX and/or MEIS members bind DNA cooperatively with HOX family members *in vitro,* which represents one mechanism that confers specificity [5,6].

Although individual homeobox genes display unique expression patterns and specific biological functions, homeobox gene families can be distinguished by certain general features of their expression patterns and functional properties, as well as by their sequence similarities. For example, HOX gene expression is generally restricted to undifferentiated and/or proliferative cells during embryogenesis in patterns that reflect their biological functions and that specify positional information [7]. Other homeobox genes have spatial and temporal expression patterns that are consistent with roles in regulating epithelial-mesenchymal interactions that are required for tissue patterning during embryogenesis [8]. By contrast, the tissue-specific expression patterns of other classes of homeobox genes in differentiated adult tissues are consistent with their functions as positive effectors of differentiation and homeostasis [9].

1.2 *Engrailed* homeobox gene family

1.2.1 *Engrailed* genes represent one of the smaller families of homeobox genes

The *Engrailed* genes are widely regarded as developmental genes. They are involved in pattern formation, neurogenesis, and neuronal differentiation [10,11,12]. In *Drosophila*, the gene *engrailed* (*en*), which belongs to the segment polarity class of genes, was first discovered as a spontaneously occurring mutation, which led to a homeotic transformation [13]. Anterior transformations and cell death occurs in the posterior

compartment of each segment in *en* mutants, and most mutants die as larvae with severely affected segmentation patterns [14]. *Drosophila en* is involved in regulating a number of key patterning processes, including segmentation of the epidermis and neurogenesis, and is an integral member of the highly complex cascade of developmental cues, which results in a fully developed fruit fly [13].

The high degree of conservation among the *Engrailed* gene family led to the rapid cloning of homologs from several species [15]. The ancestral *Drosophila en* gene has undergone gene duplication to generate two *Engrailed* orthologs in vertebrates. However, genome duplication led to four *Engrailed* orthologs in *Zebrafish,* based on their homology to the mouse *En-1* and *En-2* genes [16].

1.2.2 Molecular structure and properties of Engrailed proteins

Although individual members of homeoprotein families often share little sequence similarity outside of the homeobox, particular protein families such as Engrailed have additional conserved domains, which contribute to their distinct functional properties. Comparison of Engrailed homologs identified across several species revealed that all En proteins have short stretches of conserved regions outside of the homeodomain [17]. There are five distinct subregions within Engrailed proteins, designated EH1–5 for Engrailed homology regions, where 1 through 5 refers to their N- to C-terminal positions [18] (Fig. 1.2). The EH4 domain is the largest and most conserved region and encompasses the homeodomain. The other domains are involved in protein-protein interactions. The EH1 and EH5 domains play a role in active repression of transcription where the repressor function of EH1 *in vivo* is dependent upon its association with Groucho [19].

EH3 was defined as a region containing basic amino acids and whose primary sequence was the least conserved across species, although it is well conserved within each vertebrate class. In contrast, EH2 is 100% conserved across all species, with the exception of the flatworm where 17 of the 19 amino acids are identical. Together, the EH2 and EH3 domains mediate interactions with PBX proteins [5]. EH2 and EH3 are similar to the hexapeptide motif and the adjacent linker region in HOX proteins,

respectively, which are also 5' of the homeodomain and are responsible for mediating cooperative DNA binding interactions with PBX homeoproteins [20]. This interaction has a significant impact on the affinity of the Engrailed proteins to DNA, can redirect them to different targets, and can determine whether they act as activators or repressors of transcription [21,22,23].

There is a conserved phosphorylation site N-terminal to EH2, which is post-translationally modified by serine/threonine kinase 2 in insects and vertebrates [24]. The phosphorylation of this site may also modulate the secretion of Engrailed protein in mammalian cells [25]. Despite being a transcription factor, which are predominantly localized in the nucleus, a small proportion of intracellular Engrailed protein (less then 5%) is actually found associated with membrane vesicles [26], becomes secreted, and is internalized by cells [24,27]. This occurs despite the protein lacking a classical secretion signal and depends on a short region in the homeodomain that is essential for nuclear export and extracellular release of the protein, in addition to the phosphorylation site [28].

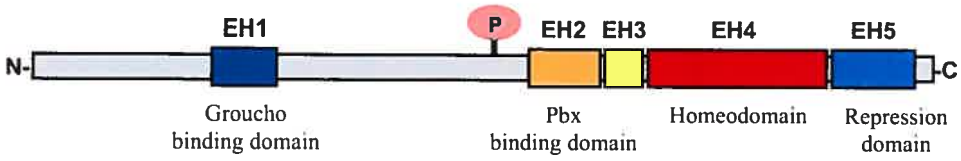

Fig. 1.2 Molecular structure of the Engrailed protein family. The five highly conserved subregions designated EH1-EH5 for Engrailed homology regions 1-5 are numerically based on their N- to C- terminal positions, and are designated as colored boxes within the full-length grey protein. The fourth and largest region, EH4, represents the homeodomain and is depicted as a red box. The letter P in the pink circle designates the conserved phosphorylation site N-terminal to EH2, which is putatively involved in modulating the secretion of Engrailed proteins in mammalian cells.

However, outside these regions of homology, En proteins share little identity across the phyla. At the protein level, the sequence differences between homologs and paralogs are significant. Whereas the homeobox domain is highly conserved, often reaching 90% amongst two different phyla, the sequence homology drops below 30% for the remaining portion of the protein. This is evident where mouse En-1 shares greater amino acid (a.a.) sequence identity with the human homolog EN1 than with the mouse En-2 paralog. Overall, mouse En-1 and human EN1 share 95% identity with each other while mouse En-1 and En-2 proteins share approximately 55% a.a. identity with each other and approx 35% a.a. identity with *Drosophila* en. Similarly, human and mouse en-2 share 90% amino acid sequence identity with each other. The En-1 class is distinguished by the proline and alanine rich regions N-terminal to EH2 in all En-1 proteins while all En-2 proteins contain a unique serine rich region. The 2 vertebrate En protein classes can be further distinguished by nine conserved amino acids C-terminal to EH5 that are specific for either the En-1 or En-2 class.

1.2.3 En-1 and En-2 in mouse development

In the mouse, *En-1* expression is first detected at the one-somite stage around 8.5 days post coitum (dpc) in cells of the anterior neural folds. *En-2* expression, which occurs in a similar region, is first detected at the five-somite stage, but does not fully overlap with *En-1* expression until approximately 8 somites have formed [29]. *En-2* continues to be expressed, along with *En-1*, within the neural plate during condensation of the first somites and in overlapping domains in the mid-hindbrain region during embryonic development [17]. At 9.5 dpc, *En-1* expression is also detected in a rostral-to-caudal pattern in two ventrolateral stripes along the hindbrain and spinal cord, in the dermomyatome of the somites, as well as in the ventral ectoderm of the limb buds, and additional *En-1* expression is later detected in the compact cells of the somite-derived sclerotome [30]. Indicative of their domains of expression during embryonic development, En-1 and En-2 are required for midbrain and cerebellum development and En-1 also plays a crucial role in dorsal/ventral patterning of the limbs and skeleton.

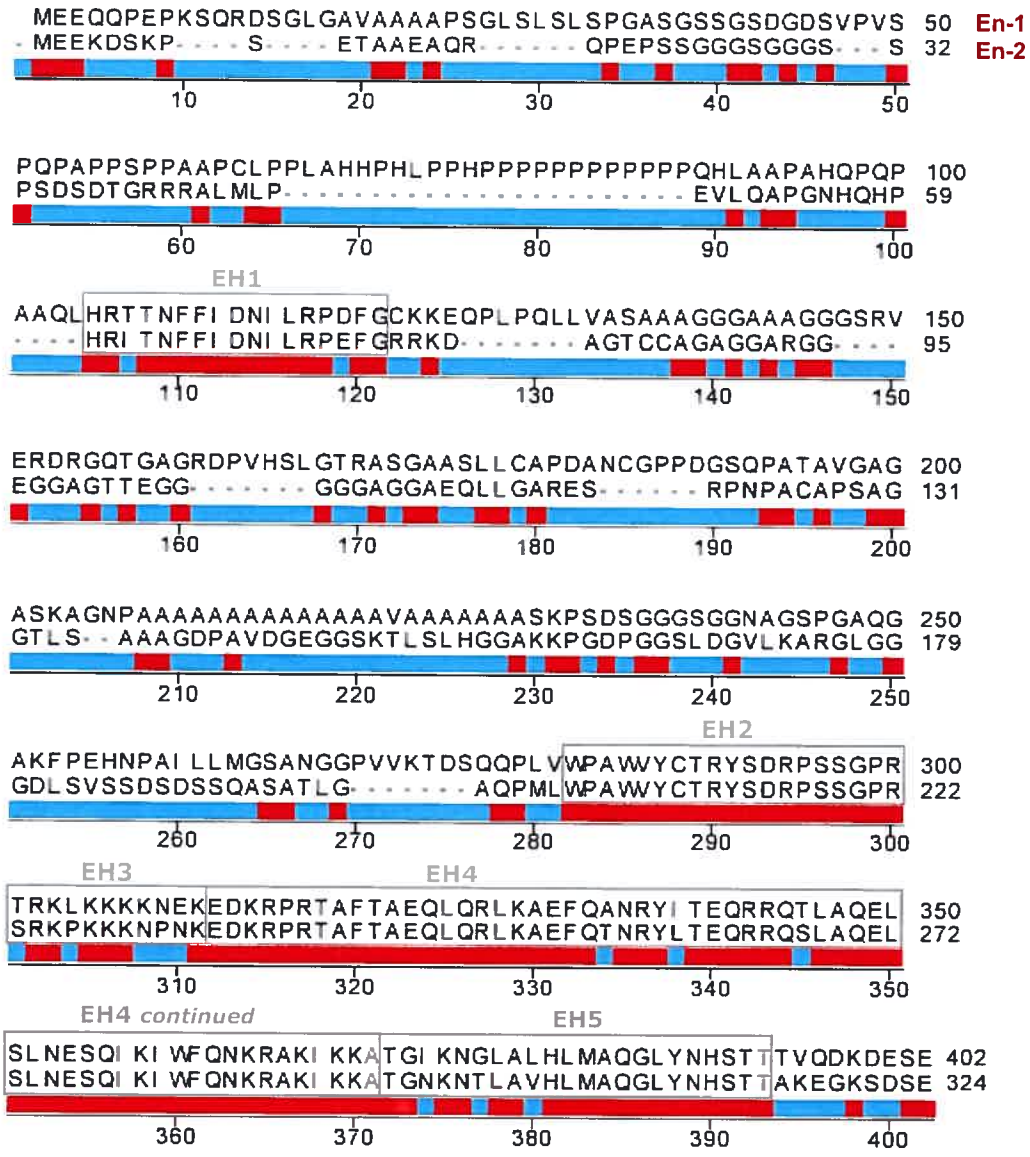

Fig. 1.3 Amino acid sequence comparison of mouse En-1 and En-2. The a.a. sequences are aligned and the five highly conserved regions EH1-5 are indicated and enclosed in grey boxes. Orange regions underneath the aligned sequences designate perfect matches while blue regions represent non-conserved a.a. Gaps introduced for alignment of the shorter 324 a.a. En-2 sequence with the 401 a.a. En-1 sequence are indicated by grey dashes.

In the adult mouse, both *En-1* and *En-2* are coordinately expressed in groups of motor nuclei in the pons region and substantia nigra that are involved in motor control. In addition, *En-2* is uniquely expressed in the granule and molecular layers of the cerebellar cells while *En-1* is uniquely expressed in the postnatal limbs [31]. Their continuing expression in the nervous system into adulthood is thought to reflect an additional function in maintaining the integrity of the central nervous system [13]. The emergence of two *En* paralogs during evolution suggests that each gene likely serves a unique role by regulating the expression of distinct target genes in specific cell types. This notion is substantiated by the finding that unique phenotypes are associated with the disruption of individual En genes in mice.

Mice homozygous for *En-1* mutations die within a day of birth and have multiple abnormalities. The *En-1* $^{hd/hd}$ (*hd*, homeobox deletion) mutants have a striking absence of the mid-hindbrain tissue that is apparent from 9.5 dpc onwards that results in the loss of the third and fourth cranial nerves and most of the cerebellum and colliculi [31]. In addition, *En-1* mutants have abnormally shaped forelimbs and exhibit skeletal defects in the 13th rib and sternum. Mutant forelimb paws are grossly deformed showing occasional ectopic ventral digits, truncations, and fusion of the digits, splaying outwards of the digits, supernumerary digits and a delay in ossification of the digits. The newborn mutants exhibited truncated sternums and a delay in ossification, as well as misalignment of the ribs and abnormal sternum ossification patterns. At birth, these mutants were readily distinguishable due to their forelimb abnormalities and by 12 hours after birth, it was obvious that these mutants were not feeding as evidenced by the lack of milk in their stomachs [31]. The mutants could move their jaws and limbs so the lack of feeding was attributed to a probable lack of appropriate innervation from the CNS for feeding due to the deletion of brain tissue.

Phenotypic studies initially focused on newborns because *En-1* $^{hd/hd}$ mutants die shortly after birth but subsequent analysis of viable mice homozygous for two other *En-1* mutant alleles, *En-1* $^{dki/dki}$ (*Drosophila en* knock-in) and *En-1* $^{2ki/2ki}$ (*En-2* knock-in), and of rare *En-1* $^{hd/hd}$ mice surviving to three weeks of age, showed the functional

importance of En-1 in specifying postnatal ventral limb structures. Around 3-4 weeks of age, the above mutant limbs develop dorsally restricted hyperpigmentation and nail-like differentiation on the ventral epidermis, ectopic ventral hairs, and occasional ectopic ventral digits that emanate from the base of the proximal mutant paw pad [32]. The *En-1* mutant brain defect provides evidence that patterning of the nervous system in mammals involves a phase of regionally controlled proliferation of cell precursors [33]. Mice lacking *En-1* function have a loss of midbrain and cerebellar structures that derive from the *En-1*-expressing brain region, suggesting that En-1 is required for the specification, survival, and differentiation of these neural precursors [31]. In contrast, loss of En-1 function in the ventral ectoderm of the developing limb does not lead to loss of *En-1*-expressing ectodermal cells, but instead results in an alteration of ventral ectoderm and mesoderm cell fate and limb patterning [32].

Mice homozygous for *En-2* mutations are viable but exhibit a distinct cerebellar phenotype where the mutant cerebellum exhibits a one-third reduction of the normal size and displays a specific alteration in the folding pattern [34]. The *En-2* mutant phenotype is milder than the *En-1* mutant phenotype and is restricted to the brain. During postnatal development of the cerebellum, En-2 is required for the production of some of the cerebellar cell precursors and for patterning and fusing of the fissures [34]. It is thought that the milder mid-hindbrain phenotype seen in *En-2* mutants is due to the fact that En-1 and En-2 share partially redundant functions in the brain where *En-1* is expressed earlier than *En-2* in the cells that will eventually express *En-2*.

1.2.4 Conservation of protein function among the Engrailed family

Despite the large sequence differences within the Engrailed family outside of the five conserved regions, the biochemical conservation of these genes is striking. Substituting mouse *En-2* coding sequences in place of mouse *En-1* coding sequences by gene targeting led to viable and fertile animals with a complete rescue of the brain defects, skeletal abnormalities and the embryonic limb patterning of the otherwise lethal *En-1* null mutant [17]. *En-1* replaced with *En-2* revealed that the two En proteins in the mouse have retained common biochemical functions throughout evolution since mouse En-2

can substitute for En-1, both in the neural tube where it is normally expressed as well as in regions such as the limbs, that normally express only *En-1* [17]. The complete rescue of the newborn *En-1* mutant phenotype suggested that the two paralogs were redundant and almost functionally equivalent. The main functional differences revealed by the null mutants for the two paralogs were thought to arise from the divergence in temporal and spatial expression rather than through divergence in biochemical function [35,34,31]. Accordingly, the more severe deletion of mid-hindbrain structures in *En-1* null mutants, in comparison to *En-2* null mutants, was thought to be due to the lack of En function in the anterior neural folds between the one- and eight-somite stages, before *En-2* is expressed.

However, when surviving *En-1* $^{2ki/2ki}$ adult mutants were analyzed, it became evident that although En-2 can rescue the embryonic limb defects of *En-1* $^{hd/hd}$ mutants, it cannot rescue the postnatal patterning defects of *En-1* $^{hd/hd}$ mutant limbs. *En-1* $^{2ki/2ki}$ mice developed the hyperpigmentation and ventral nail-like structures of *En-1* $^{hd/hd}$ mutants 3-4 weeks after birth [36]. This is a direct demonstration in mammals that the regulation of essential steps during embryogenesis has been conserved by the two paralogs, but the fact that En-2 cannot rescue the postnatal limb abnormalities shows that En-1 has acquired some novel functions during evolution above those that are seemingly redundant with En-2.

The same replacement experiment was repeated by replacing the mouse *En-1* coding sequences with *Drosophila en* coding sequences and clearly demonstrated the functional homology across phyla, when the ortholog from an invertebrate was able to substitute for a mammalian gene throughout the development of a highly ordered structure, the brain [36]. The resulting mice were viable and fertile, further demonstrating the biochemical conservation over hundreds of millions of years of evolution. Mice expressing *Drosophila en* in place of *En-1* results in a near complete rescue of the lethal *En-1* mutant brain defect and most skeletal abnormalities, revealing a common underlying molecular mechanism to their diverse developmental activities. In contrast, surviving adult *En-1* $^{dki/dki}$ mice demonstrated that expression of *Drosophila en* cannot

functionally replace En-1 in the dorsal/ventral patterning of the limbs during either embryonic or postnatal development [36].

Although neither En-2 nor en are capable of rescuing the postnatal limb abnormalities that develop in rare *En-1* $^{hd/hd}$ mutants that survive, these studies demonstrate that the biochemical activity utilized in mouse to mediate brain development has been retained by Engrailed proteins across the phyla, and indicate that during evolution vertebrate En proteins have acquired two unique functions during embryonic and postnatal limb development and that only En-1 can carry out the latter [36].

All En proteins share 5 conserved domains (EH1-EH5), and coding sequences in the non-conserved regions of *Drosophila en* could have evolved a functional domain required for vertebrate limb development, which became further specialized in *En-1* after the second *En* gene (*En-2*) was formed by duplication [18]. The functional differences between the 3 En proteins may reflect the inability of en and En-2 to interact with the full repertoire of En-1 accessory proteins, possibly resulting in altered DNA binding affinities for selective targets.

1.2.5 Engrailed is a target of the Wnt pathway

Several upstream regulators that activate, repress or maintain *engrailed* expression in the developing *Drosophila* embryo have been identified, and similarly, several direct and indirect targets of engrailed regulation have been identified in *Drosophila* but very few mammalian regulators or targets of *Engrailed-1* and/or *Engrailed-2* have been identified. It has been shown, however, that the maintenance of *En-1* expression during embryonic development by the Wnt signaling pathway has been conserved from flies to mice [37]. Genetic analysis of *Drosophila* development indicates that *wingless* (*Wnt-1* homolog) signaling in the epidermis is required for the maintenance of *engrailed* expression in adjacent cells [38]. In *Wnt-1*$^{-/-}$ mouse embryos, initial *En-1* and *En-2* expression is normal, but subsequent expression of both are lost [29]. In addition, compound *En-1* x *En-2* mutants have a similar phenotype to that of *Wnt-1*$^{-/-}$ mutants where the midbrain and anterior hindbrain fail to develop [39]. The functional link between En-1 and Wnt-1 was

best shown when *En-1* expression, under the control of the *Wnt-1* enhancer, was sufficient to rescue the early midbrain and anterior hindbrain phenotypes in *Wnt-1*$^{-/-}$ embryos [37]. *Wnt-1* encodes a secreted growth factor that initiates a signaling cascade, which results in transcriptional activation mediated by *ß*-catenin/Tcf complexes [40].

1.3 Mouse Mammary Gland Development

The mammary gland forms as an appendage of the skin and has its evolutionary origin in skin glands [41]. All mammary glands reside just underneath the skin but the number and location vary among different classes of mammals. For example, only one pair of mammary glands develops in the thoracic region in humans whereas mice possess five ventral pairs of mammary glands (Fig. 1.4). The mammary gland consists of two main components. The epithelial component or parenchyma refers to the epithelial system composed of ducts and milk-producing alveolar cells within the gland. This extensive ductal system is embedded in the surrounding stromal component, which refers to the fatty connective tissue of the mammary gland that supports the epithelial component. Adipocytes account for the majority of cells in the stromal compartment, but fibroblasts, cells of the hematopoietic system, blood vessels and neurons also reside in the fat pad [41]. The epithelial cells form the branched system of ducts that channel into a main primary duct, which opens up to the body surface through the nipple. A large proportion of epithelial cells within the mammary gland are luminal secretory cells, which undergo functional differentiation during pregnancy to produce milk, which is then secreted into the inner lumen of the ducts. Basal myoepithelial cells surround the ductal system and these contractile cells facilitate the delivery of milk from the milk secreting cells to the nipple during lactation.

Unlike most mammalian organs, which develop primarily embryonically, the mammary gland is established during fetal development but the majority of expansion and development occurs postnatally [42]. The structural and functional development of the gland involves the influence of hormones and growth factors like estrogen, progesterone and prolactin and development of the gland itself can be divided into roughly six distinct stages; embryonic, postnatal, puberty, pregnancy, lactation and involution (Fig. 1.5).

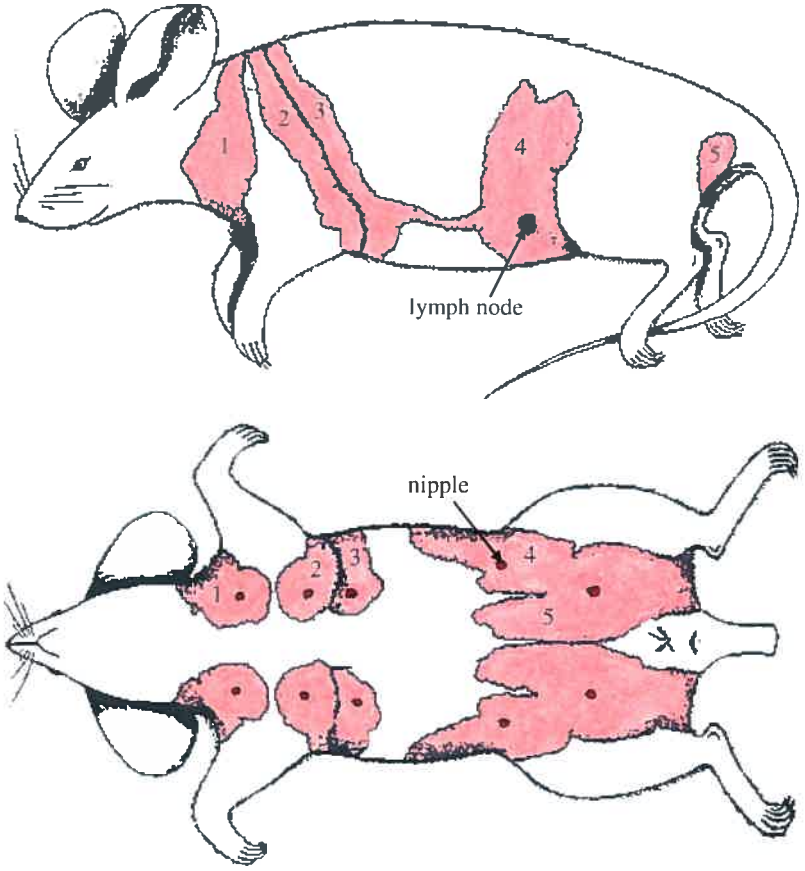

Fig. 1.4 Location of the five pairs of visible ventral nipples and corresponding mammary glands just under the skin in the female mouse. The nipples are shown in red and the mammary fat pad is shown in pink (the mammary epithelial tree within the fatty stroma is not depicted here). The adipose stroma of the gland provides a frame of support as well as a substrate within which the mammary epithelial tree can grow and function. The first 3 mammary glands are in the thoracic region while the 4^{th} and 5^{th} mammary glands are in the abdominal region. The 4^{th} (inguinal) mammary gland is the most accessible and thus, is the gland of choice in most experimental manipulations. The 4^{th} mammary gland also contains a central lymph node (shown in purple), which serves as a convenient reference point. Adapted from DeOme K.B. *et al.* (1959) *J. Natl. Cancer Inst.* 78:515-525.

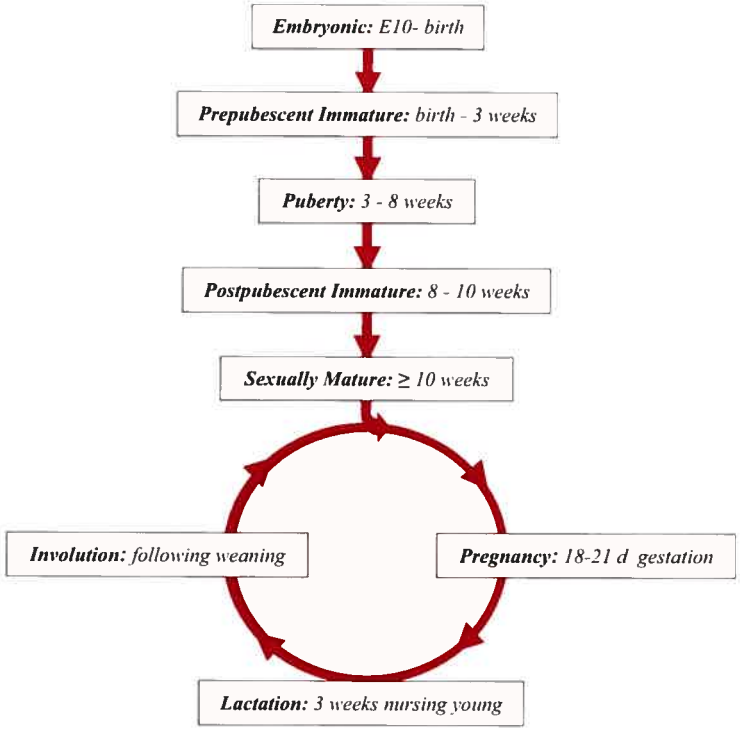

Fig. 1.5 Distinct stages within mammary gland development. The linear portion of the diagram represents the proliferative expansion and morphogenesis that takes place in the embryonic and virgin female. The circular portion of the diagram represents the cycle of proliferative expansion and morphogenesis, functional differentiation and eventual remodeling accompanied with each pregnancy. Adapted from Lewis M.T. (2000) *Breast Cancer Res.* 2:158-169.

1.3.1 Formation and differentiation of the embryonic mammary gland

The mammary epithelium is an ectodermal derivative and therefore, the first distinction that must be made is the differentiation of the presumptive mammary epithelium from tissue that can also differentiate to form skin, hair follicles or other ectodermally derived structures. The future mammary anlage begins to differentiate from the ectoderm at day 10 of gestation (E10) with the establishment of the mammary streaks, two lines of

epidermally derived thickened epithelium that extend from the anterior to the posterior limb bud, symmetrically displaced off the ventral midline of the embryo. These streaks represent the first morphological evidence of mammary pattern formation and differentiation before sexual differentiation of the gonads. The mammary anlage is first visible as small placodes which define the nipple region around E11.5 and then appears as five pairs of small epithelial buds on the ventral side of the embryo on E12 that enlarge to form bulb-like structures by E14. During embryonic development the mammary epithelium is associated with two types of mesenchyme; the mammary mesenchyme that is directly attached to the epidermal bud and the future mammary fat pad that is located below the epithelial bud in the deeper mesenchyme and consists of preadipocytes (Fig. 1.6). The mammary mesenchyme consists of several layers of concentrically organized fibroblasts, which surround the epithelial bud and are more densely packed than the dermal cells while the mammary fat pad appears at E14 as undifferentiated mesenchyme and is required for future mammary epithelial morphogenesis.

The cell identity of mammary epithelium is clearly established as early as E12.5 as shown by the capability of these epithelial buds at this stage to generate a ductal tree when transplanted into a cleared fat pad of a syngeneic host. Mammary rudiments are formed in both sexes but the sexual phenotype of the mammary gland is determined between E13 and E14. During this period the gland displays responsiveness to steroid hormones and is influenced by signals from the surrounding mesenchyme. In male embryos testosterone acts on the dense mesenchyme surrounding the epithelial mammary rudiment, which results in the subsequent detachment of the gland from the epidermis. In females the mammary buds continue to grow and reciprocal interactions between the mammary bud and the mammary mesenchyme take place, which are crucial for elongation of the mammary bud. As the mammary bud elongates to form a primary sprout, it reaches and invades the second mesenchyme, the future mammary fat pad by E16. The primary sprout then undergoes a small amount of branching morphogenesis, which leads to the rudimentary ductal epithelial tree of the neonatal female.

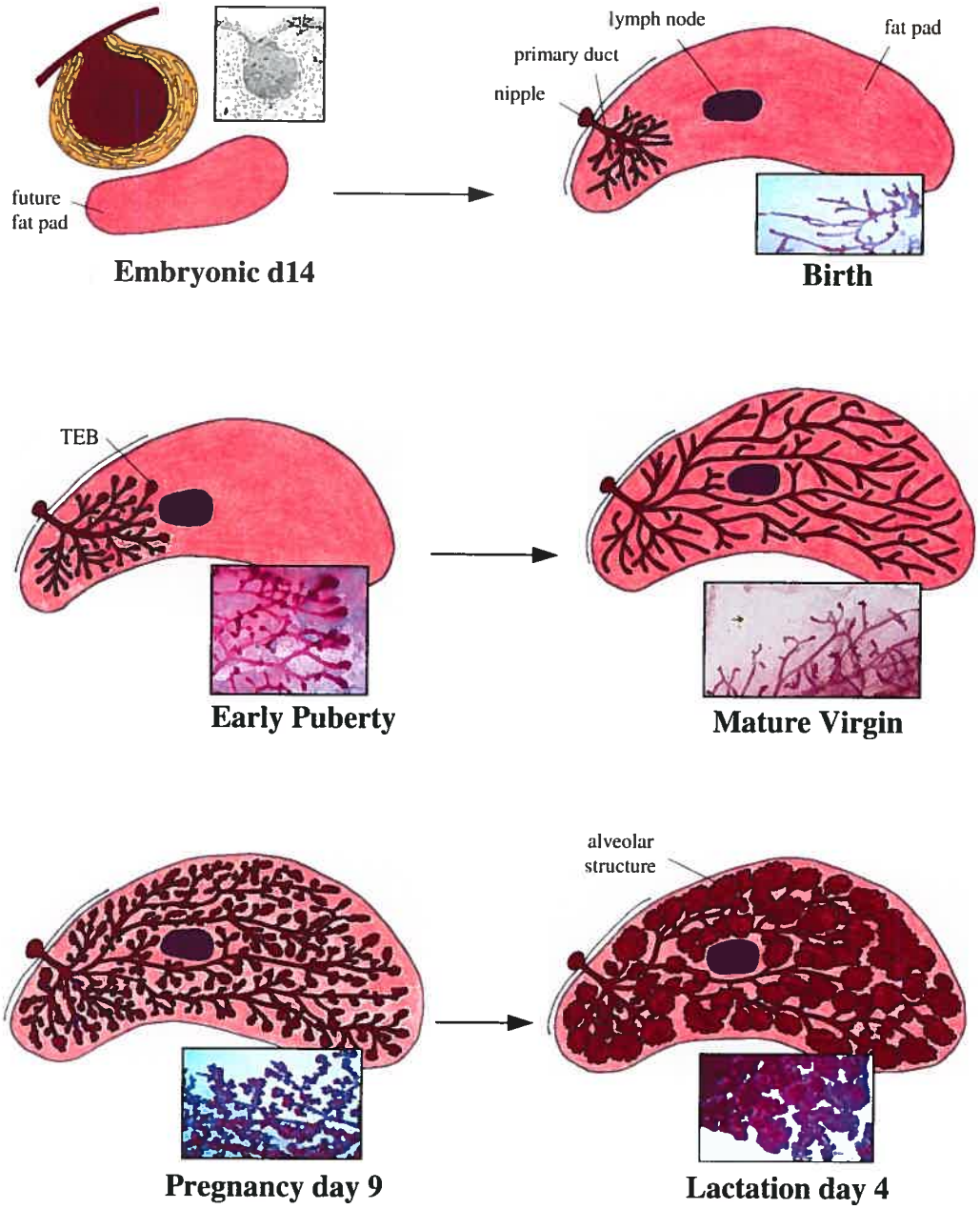

Fig. 1.6 The inguinal (#4) mouse mammary gland depicted at different developmental stages. The adipose stroma of the mammary fat pad is shown in pink while the mammary epithelial system is depicted in red. The central lymph node, which serves as a convenient reference point to evaluate ductal outgrowth in the #4 mammary gland, is shown in purple. The embryonic mammary epithelial bud is surrounded by mammary

mesenchyme (yellow) while the future mammary fat pad lies below (pink). In the newborn mouse, the few rudimentary ducts emanate from a central primary duct that is connected to the nipple. At this stage, the rudimentary ductal system is proximal to the nipple and occupies little space in the mammary fat pad. Rapid and invasive ductal elongation and branching commence with puberty and terminal end buds (TEBs) start to become visible. The ductal pattern of the mammary epithelial tree is created by the bifurcation and penetration of these TEBs through the underlying stromal fat pad. The growing ductal system reaches the central lymph node around 4 ½ weeks, and continues growing until the ducts have reached the peripheral limits of the mammary fat pad around 6-8 weeks of age. Additional branching occurs in the postpubescent immature virgin and at 10-12 weeks the mammary gland reaches sexual maturity and ductal elongation ceases and TEBs regress to leave a branched system of ducts. Proliferation of mammary secretory epithelium during pregnancy leads to the formation of lobuloalveolar structures, which resemble grape-like clusters. At parturition the mammary fat pad is completely filled with secretory alveolar structures and milk proteins are secreted in large amounts by the secretory epithelial cells.

1.3.2 Relatively growth quiescent prepubescent mammary gland development

At birth, the mammary epithelial tree consists of 15-20 rudimentary branches emanating from a central primary duct, which is connected to the nipple (Fig 1.6). The rudimentary system of small ducts present in the newborn female occupies only a small portion of the mammary fat pad in the vicinity of the nipple. During the first three weeks of postnatal life, before the onset of gonadal hormone secretion, the mammary ducts are in fact slowly elongating and branching into the underlying fatty stroma at a rate that is in pace with the overall growth of the entire animal.

1.3.3 Proliferation and morphogenesis in pubescent mammary gland development

Functional development of the mammary gland proceeds in distinct stages that mostly coincide with the hormonal influences of puberty and pregnancy. Around 3-4 weeks of age, ovarian hormones stimulate accelerated and invasive ductal extension and

branching whereby the growth rate of the epithelial ductal system now exceeds the overall growth rate of the animal. Estrogen, along with other hormones, plays a critical role in the expansion and morphogenesis of the growing ductal system.

The onset of puberty also coincides with the appearance of large club-shaped TEBs, highly proliferative and active structures found at the tips of the growing ductal branches. It is during the early pubertal developmental stage (3-7 weeks of age) when the TEBs are most prominent (Fig 1.6). These TEBs are influenced by systemic steroid hormones and aid the ducts in linear growth as well as the regulation of branching patterns. The ductal pattern of the mammary epithelial tree is created by the bifurcation and penetrating extension of these TEBs through the underlying stromal fat pad. The TEB is a specialized structure, comprising a solid mass of epithelial cells, which is composed of two distinct relatively undifferentiated cell types. Actively proliferating cap cells make up the outermost layer of the end bud and interact with the surrounding stroma though a thin basal lamina as the subtending duct is formed while body cells (about 6-10 layers thick), fill the interior of the end bud. As the ducts elongate and the TEBs move forward, it is thought that the inner body cells are the precursors of luminal epithelial cells and give rise to the inner luminal epithelial cell layer of the subtending duct and the outer cap cells are the precursors of myoepithelial cells and give rise to the outer myoepithelial cell layer of the newly formed portion of the duct. Myoepithelial cells are contractile cells that form a sleeve around the primary ducts and become discontinuous around secondary and tertiary ducts and the TEBs themselves. Luminal epithelial cells are generally used to refer to the non-myoepithelial component of the mammary epithelium system and these cells line the lumen, the space within the centre of the mammary ducts.

Dichotomous branching of the growing ductal system occurs at the site of TEBs as they penetrate the underlying fat pad while monopodial branching of ducts occurs by budding from the body of the existing ducts. Ductal morphogenesis and inner lumen formation is accomplished by a highly regulated process of both cell proliferation and death in the TEB. It has been demonstrated that apoptosis is an important mechanism in ductal

morphogenesis during puberty and occurs in the middle of the mass of body cells and adjacent developing luminal cells to generate the ductal lumen [43]. While a variety of cell types have been identified and their developmental capacities have begun to be explored, very little is known about how these cell lineages and fates are established at this pubertal stage.

The growing ductal system reaches the central lymph node around 4 ½ weeks, and continues growing until the ducts have reached the peripheral limits of the mammary fat pad around 6-8 weeks of age, at which time few TEBs remain. Additional branching occurs in the postpubescent immature virgin and at 10-12 weeks the mammary gland reaches sexual maturity and ductal elongation ceases and TEBs regress to leave a branched system of differentiated ducts. The fully differentiated mature mammary ducts can now serve as channels for milk transport during lactation and consist of a discontinuous outer lining of myoepithelial cells lined by a single layer of luminal epithelial cells. The gland becomes essentially quiescent in the mature virgin except for brief periods during the ovulation cycle and the ducts will remain relatively quiescent as long as the mature female remains a virgin.

1.3.4 Alveolar growth and secretory differentiation during pregnancy

Another phase of rapid proliferation takes place during pregnancy where 50% of the overall growth of the gland takes place from pregnancy day 12 until parturition. Hormonal changes during pregnancy initiate this cyclical phase of development which leads to a dramatic transition from a predominantly ductal morphology to a predominantly lobuloalveolar gland morphology. During pregnancy, the gland comes under the influences of estrogen, progesterone and other placental hormones. New ductal outgrowth occurs from the lateral walls of the ducts and side buds increase in number. In addition, lobuloalveolar progenitor cells located in the ducts, mostly at the terminal ends of the ducts, proliferate and undergo alveolar development to form alveolar buds which differentiate to form grape cluster-like structures containing lobuloalveolar units (Fig 1.6). This vast expansion of the mammary epithelium fills in most of the fatty stroma between the ducts. The alveoli are composed of a web like network of myoepithelial

cells on their outer surface surrounding columnar epithelial cells facing the lumen, the space in the centre of the alveoli where milk is initially secreted.

These morphological changes that the alveolar epithelium undergoes are accompanied by the development of secretory epithelial cells within the alveoli that acquire the ability to produce milk proteins by midpregnancy. The capability to produce milk proteins represents the first stage in the transition to lactogenesis but the secretion of these milk proteins is inhibited during pregnancy. At parturition, secretory function is no longer inhibited and milk proteins are secreted in large amounts by the secretory epithelial cells, which represents the second and final stage in the transition to lactogenesis. Lobuloalveolar development and proliferation occur during pregnancy and result in the complete filling of the fat pad by parturition while functional differentiation of the secretory epithelium coincides with parturition and lactation. Cell division occurs in both the alveolar and ductal cell populations throughout pregnancy and continues into the mid stage of lactation. Pregnancy terms vary slightly between 18-21 days in different mouse strains.

1.3.5 Large quantities of milk production and secretion accompany lactation

Lactation involves the production and secretion of milk by the secretory epithelial cells. Mammary epithelial proliferation continues into early lactation where it is estimated that 20% of total mammary growth occurs during the first 14 days of lactation. By the time the mother is feeding her pups, her mammary glands are packed full of secretory epithelium with little fat, the complete opposite of the situation in the virgin or non-pregnant animal. The initiation of lactation is thought to be induced by the decrease in estrogen and progesterone and several hormones such as prolactin, insulin and glucocorticoids are involved in the maintenance of lactation. While the inner luminal cells of the alveoli produce the milk, the outer contractile myoepithelial cells form a basket-like network around the secretory alveoli and these cells are responsible for squeezing the milk out of the alveoli and down the ducts and out of the nipple in response to the hormone oxytocin.

1.3.6 Apoptosis, regression and remodeling accompany involution after weaning

When lactation ceases after weaning, the mammary gland undergoes involution where the entire mammary alveolar compartment is remodeled. This remodeling elicits a dramatic change in the morphology of the mammary gland, restoring it to its pubescent state. The process of involution commences with the suspension of milk production, followed by the collapse of the mammary alveolar structures and removal of the secretory epithelial cells through programmed cell death and phagocytosis. The alveolar structures collapse likely as a result of degradation of the extracellular matrix (ECM) and apoptosis of secretory epithelial cells and surrounding myoepithelial cells. The basement membrane and most of the epithelial cells are replaced with adipose tissue so that the mammary gland changes from the epithelial rich lactational state to the epithelial sparse non-parous state. After regressing to this state following weaning of the offspring the fat pad contains only well spaced ductal structures within the adipose matrix. With each subsequent pregnancy, a new cycle of lobulo-alveolar development occurs and can be repeated several times during the life of an animal.

1.3.7 Mammary gland transplantation

An essential feature of the mouse mammary gland is the regenerative capacity of its epithelium. Any portion of the epithelial tree can be transplanted into the mammary fat pad of a syngeneic female, whose endogenous epithelial tree has been removed, and reproduce a complete and functional mammary gland [44]. In a 3 week old female mouse, the epithelial ductal system occupies little space in the mammary fat pad and is confined to the most proximal portion of fat pad connected to, and in the vicinity of, the nipple (Fig. 1.6). This area containing the prepubertal epithelial ductal system can easily be surgically removed to generate a 'cleared' mammary fat pad into which cells or tissue fragments from another female can be transplanted where it can develop under the influences of a wildtype hormonal environment. In addition, endogenous epithelium can be left intact, providing identical mammary fat pad conditions to enable *in situ* comparisons between wildtype and transplanted epithelium.

The deletion of a gene in the mouse genome can sometimes lead to the disruption of the normal development and function of more than one organ. The absence of some genes can influence ovarian function, which would obscure assigning direct or indirect effects on mammary gland development. Moreover, some homozygous deletions are not viable and postnatal mammary gland development cannot be studied directly. The above issues can be overcome by transplanting mammary epithelial cells from such mutants into a wildtype syngeneic host.

The ability to transplant portions of the mammary epithelium into the cleared fat pad of a syngeneic host allows one to examine both the morphogenic and tumorigenic capabilities of that particular mammary epithelium. Cell populations removed from the mammary gland and mammary epithelial cells grown *in vitro* can also be transplanted. The 4^{th} (inguinal) mammary gland is the most accessible and is the gland of choice in most experimental manipulations (Fig. 1.4).

The ability to delete genes from the mouse genome, in conjunction with tissue transplants to evaluate their physiologic role, has led to the identification of several genes involved in mammary gland development. Experimental manipulation of mammary tissue from wildtype and knockout mice has shed light on distinct signaling networks activated by systemic hormones that induce or are involved in mammopoiesis while some of these signals act through reciprocal interactions between the mammary epithelium and the adipose stroma. The mammary transplantation method has been employed for decades to examine fundamental questions in mammary morphogenesis, senescence and tumorigenesis. Such experiments examine the role of genes in the context of a normal tissue environment with an endogenous hormonal milieu and can determine the potential role of oncogenes in preneoplastic and neoplastic transformations and the demonstration of oncogenic potential of unknown genes.

Daniel and colleagues used serial transplantion studies to investigate the pattern of senescence in mammary cells and demonstrated that the proliferation potential of normal mammary cells declined with serial transplantation and was lost after 5-6 serial

transplants [45]. The number of prior cell divisions rather than chronological age of the transplant donor was the most important determinant for the onset of senescence or lack of division potential in the mammary gland. Moreover, they also demonstrated that mouse mammary preneoplasias did not senesce and were essentially immortal populations.

1.3.8 Adult mammary epithelial stem cells

The observation that mammary epithelial fragments or cells from adult mammary glands can generate a fully differentiated mammary tree capable of lactating when transplanted into the cleared fat pad of a prepubertal mouse led to the speculation that the mammary epithelium must contain multipotential stem cells. Mammary epithelial stem cells (MSCs) have been the focus of much recent interest because they are also likely the cells of origin of breast tumors. Some have suggested that the notion of breast stem cells as a resident cell population in the adult breast is an oversimplified view and that perhaps a selective cell population has MSC-like ability, possibly as a result of its interactions with its microenvironment or stem cell niche. No definitive identification has been made of an adult MSC yet experimental evidence supports the notion that such an elusive stem cell exists and would give rise to the distinct mammary epithelial cell lineages within the mammary gland and models of how these lineages might develop have been proposed (Fig. 1.7). Currently, at least four distinct mammary cell populations are thought to exist. One is the multipotent MSC that is capable of self-renewal and recapitulating the entire mammary ductal system, and these MSCs would gives rise to committed epithelial precursor cells (EPCs). The progeny of these EPCs give rise to two pre-committed epithelial progenitors that become restricted to a ductal or alveolar fate [41]. The ductal precursor cells (DPs) form myoepithelial cells and luminal cells, the two cell types that exist in ductal structures. During pregnancy, the alveoli are thought to be generated from alveolar precursors (APs), which give rise to myoepithelial cells and luminal cells.

The pre-committed epithelial progenitors and the stem cells that are capable of giving rise to both cell types, are thought to exist throughout the entire mammary epithelial tree but may be enriched in TEBs [46]. During the rapid expansion of the ductal system during

puberty where the body cells and cap cells proliferate and differentiate to generate new sections of the growing duct behind the active TEBs invading the underlying fat pad, stem cell-like activity is thought to reside in the TEBs. Since some of the cap cells can be seen to migrate into the body cell mass, the cap cells are the proposed stem cells and symmetric cell divisions of the cap cells within the TEBs as they move through the fat pad are thought to deposit MSCs in the growing ducts. In the mature virgin mammary gland, the location of putative MSCs is thought to be suprabasal, at the base of the luminal epithelial cell layer, adjacent to the myoepithelium and not contacting the lumen or the basement membrane [47]. This proposed stem cell niche would contain undifferentiated MSCs that do not express markers of either myoepithelial or luminal epithelial cells. Divisions of this MSC would generate progenitor cells that, at least in its early stages, would be very difficult to distinguish from its parental MSC until it acquires both myoepithelial and luminal lineage markers before becoming committed to either the luminal or myoepithelial lineage [41]. In the proposed model, the ductal system contains multipotent stem cells and committed ductal and luminal precursor cells nestled throughout the remodeled ductal system after involution (Fig. 1.7).

MSCs are conceivably the only cell population that would possess the replicative potential that would be needed to maintain the routine tissue renewal, the massive expansion in epithelial tissue that accompanies pregnancy and the cyclical process of subsequent pregnancies in the mammary gland. Such candidate stem cells would be the type that are quiescent until responding to physiological cues. In non-parous animals, a less expansive but similar process occurs as the estrus cycle progresses, as evidenced by the slight expansion and regression of the alveolar buds in response to the cyclic hormonal influences. Conversely, MSCs in the postnatal mouse mammary gland may also be involved in the replacement of luminal epithelial cells that are shed from the inner lining of the ducts into the lumen during routine cell turnover. Evidence to support this occurrence during lactation is provided by the fact that epithelial cells can be recovered from milk. Cells that are shed into the lumen of the alveolar and ductal systems must need to be continually replaced in order for the mammary epithelial tree to maintain its structure and integrity. Candidate stem cells that would participate in this

process would need to be continually active. Thus, there are two distinct potential roles for MSCs in the adult mammary epithelium but whether this infers that there might be two or more distinct MSC populations, perhaps in a stem cell hierarchy, or that one MSC population has the capacity to perform several functions depending on the environmental cues it is given, remains to be elucidated.

The analysis of stem cell function can be complicated by the presence of progenitor cells that have no self-renewal capacity, unlike the parental MSC, but that undergo a population expansion to increase the number of fully differentiated cells that are ultimately produced by the original MSC division event. Retroviral tagging and transplantation experiments have provided an estimate of the number of potential stem cells in the mouse mammary epithelium that could generate epithelial structures when transplanted into a cleared fat pad at approximately 1 in 2,500 cells [48]. Since the number of mammary epithelial cells in a mature virgin mouse mammary gland has been estimated at $2–2.5x10^6$, each gland would contain roughly 1,000 stem cells [49].

Freshly isolated mammary and breast epithelial cells can be cultured in defined medium conditions where the cells form nonadherent aggregates called mammospheres [50]. These mammospheres can be disaggregated and the resulting cells that are recovered can be divided and cultivated in separate conditions to identify those that are capable of forming secondary mammospheres and those that attach to the culture dish and differentiate [50]. Employing lineage specific markers, these studies have shown that sphere forming cells share several characteristics attributed to MSCs that can be assayed *in vitro*, including a high proliferative potential and multipotentiality. These *in vitro* stem cell assays are indirect and retroactive in that the existence and the properties of a stem cell population are inferred solely by the analysis of their mature cell progeny and the only definitive way to reveal the existence of MSCs is by *in vivo* transplantation. Analagous to the relationship between long-term culture initiating cells (LTC-IC) and hematopoietic stem cells (HSC), mammospheres are not necessarily MSCs and may be progenitor cells which are still multipotential in their differentiative capacity but have a

significantly reduced self-renewal ability, or they may even represent artifacts of cell culture.

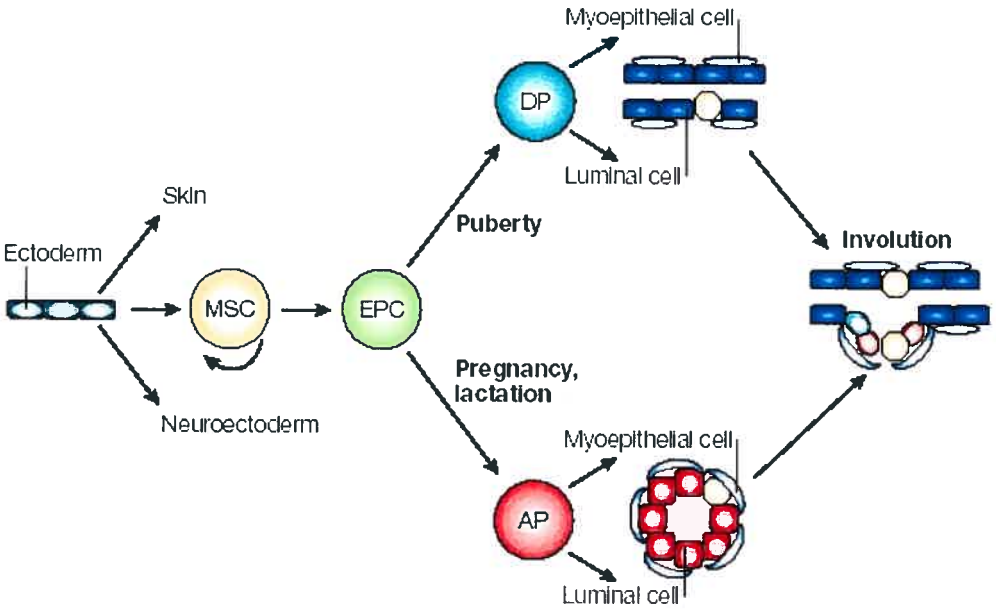

Fig. 1.7 The proposed and elusive mammary epithelial stem cell (MSC) would give rise to distinct epithelial cell lineages in the mammary gland. The mammary gland is a derivative of the ectoderm, which also gives rise to the skin and other appendages as well as the neuroectoderm. It is thought that the multipotent MSCs would give rise to epithelial precursor cells (EPCs), the progeny of which develop into either ductal or alveolar structures. Ductal precursors (DP) give rise to basal myoepithelial cells and luminal cells as the ducts are generated postnatally, particularly during puberty. The initiation of pregnancy coincides with alveolar precursor (AP) cells giving rise to basal myoepithelial and luminal epithelial cells. After lactation, the alveolar cells are subject to programmed cell death during the process of involution. A simple ductal system containing multipotent (yellow) and committed ductal (green) and luminal (orange) precursor cells persists that will develop into a fully functional epithelium in subsequent pregnancies. Figure reprinted from Hennighausen, L. and Robinson, G.W. (2005) *Nat. Rev. Mol. Cell Biol.* 6:715-725.

1.4 Role of Homeobox genes in the mammary gland

1.4.1 Homeobox genes have evolved to have roles in mammary gland development

Vertebrate homeobox genes are well known for their roles during embryonic development, and new functions are continually being identified in embryogenesis, but less attention has been devoted to investigating their potential roles in the adult vertebrate. The notion that such important regulatory proteins may exert additional functions in adult life has been supported by reports where mice lacking the function(s) of one or several homeobox genes affected adult development processes such as the solidity of hairs, the size of the prostate and the hematopoietic system [51]. These reports are slowly growing in number and suggest that complex processes in the adult mammal have evolved to employ genes whose ancestral functions are unrelated.

For example, *Hoxa10*, *Hoxa11* and *Hoxd11* are involved in the urogenital system of adult mammals during pregnancy and in addition, their expression varies in response to the estrous cycle and pregnancy [52]. Based on their important functions during the development of other organs, homeobox genes are prime candidates for being involved during the postnatal development of the mammary gland. Immunohistochemical studies and screens based on polymerase chain reaction (PCR) analysis first detected homeobox gene expression in several mammary epithelial cell lines and in the normal developing mammary gland [53]. Several Hox members appear to exhibit developmentally regulated expression patterns during postnatal development and their expression is found in either the mammary epithelium or in the periductal stroma or both. During embryogenesis, *Hoxb9* and *Hoxd9* are expressed in the condensed mammary mesenchyme surrounding the mammary epithelial bud, suggesting roles in mesenchymal-epithelial interactions during the formation of the rudimentary mammary gland [54]. *Hoxd9* is expressed throughout postnatal development in the mammary gland and is expressed at high levels in the ductal epithelium and periductal fibroblasts in the virgin. *Hoxd10* is expressed in both the mammary epithelium and stroma during lobular differentiation before lactation and targeted disruption of *Hoxd10* leads to lactation failure in homozygous mutants [3].

Mammary glands of female compound mutants lacking the function of Hoxa9, Hoxb9 and Hoxd9 do not undergo the proper development and differentiation during pregnancy and after pregnancy, leading to a strong deficit in the ability to produce milk [54]. Consistent with roles in regulating cellular proliferation in the adult mammary gland, the extensive branching process in the compound mutants during pregnancy was significantly reduced, leading to hypoplasia of the normally proliferative and expansive pregnant mammary gland. Since these females have an abnormal lactational capacity, they are subsequently unable to properly feed their young.

Homeobox genes *Msx-1* and *Msx-2* are expressed in reciprocal tissue compartments in the postnatal mammary gland; *Msx-1* is expressed in the mammary epithelium while *Msx-2* is expressed in the periductal stroma. Interestingly, *Msx-2* was also found to be responsive to estrogen and may have a potential role in mediating hormone responses [55]. Analysis of Msx function in the postnatal mammary gland has been hindered by the fact that both homozygous mutations in *Msx-1* and *Msx-2* are embryonically lethal.

Given the large number of homeobox genes in the genome and the small number whose potential involvement in mammary gland development has been studied, much remains to be answered with respect to homeobox gene expression and function in mammary tissue.

1.5 Role of Homeobox genes in cancer

It is widely accepted that many of the molecular pathways that underlie carcinogenesis represent aberrations of the normal processes that control embryogenic development. There are many examples in which the aberrant expression of homeobox genes that normally regulate growth and development have been implicated in carcinogenesis, making them an ideal candidate gene family to study the relationship between embryogenesis and oncogenesis. Despite numerous reports of deregulated expression of homeobox genes in cancer, few studies have established direct causal links or functional roles above a mere correlative expression profile, or whether their altered expression promotes carcinogenesis. It was originally thought that homeobox genes were

transcriptional activators that promote oncogenesis through their aberrant upregulation in carcinoma cells. Although deregulated homeobox gene expression was originally associated with oncogenic activities, it is now apparent that homeobox genes might be lost [56] as well as gained in cancer, and their corresponding activities might be tumor suppressing as well as tumor promoting. Most cases of deregulated homeobox gene expression in cancer conform to a simple rule: those homeobox genes that are normally expressed during development in undifferentiated cells are upregulated in cancer, whereas those that are normally expressed in adulthood and/or in differentiated tissues are downregulated in cancer [7].

1.5.1 Homeobox genes are implicated in cancer

In addition to documenting homeobox gene expression in the normal mammary gland, *HOX* gene expression was also detected in several primary breast tumor samples [53]. There are many cases in which *HOX* genes are re-expressed in tumors and carcinoma cell lines that are derived from tissues in which *HOX* genes are normally expressed during development. There are a few examples however, where the cancerous tissue that ectopically expresses a homeobox gene is not a derivative of a tissue where that particular homeobox gene is normally expressed during development. For example, *PAX5* is expressed in medulloblastoma, but it is not normally expressed in the cerebellum from which this tumor is derived [57]. The other case lies in homeobox genes that are downregulated in tumorigenic cells that are derived from tissues in which that particular homeobox gene is normally expressed during the differentiation of these cells. The loss of expression of *NKX3.1* in prostate cancer fits into this category [58]. Homeobox genes that are expressed in carcinoma are typically those whose normal expression pattern is restricted to undifferentiated or proliferative cells. In contrast, homeobox genes that are downregulated or lost in carcinoma are normally expressed in fully differentiated tissues.

1.5.2 Misexpression of homeobox genes in carcinoma

In support of a causal role in carcinogenesis, gain of function of certain homeobox genes has been shown to promote the transformation of cells in culture. Expression of *HOX* genes in nontransformed fibroblast cell lines leads to an increase in proliferation, other hallmarks of oncogenic transformation in cell culture and tumor growth when cells are injected into nude mice [59]. For some homeoproteins, it is thought that the wildtype, rather than a mutant form of the protein, produces these oncogenic activities. This implies that the oncogenic activity of some homeoproteins is not due to new or alternative properties, but is a result of their normal functions being carried out in the wrong cellular context. For example, *MSX* genes have been shown to inhibit differentiation during development and purportedly upregulate *cyclin D1*. Cyclin D1 is involved in regulating the G1-S transition so it has been proposed that *MSX* genes block terminal differentiation during embryogenesis by preventing cells from exiting the cell cycle [60]. The oncogenic properties of *MSX* genes in cancer cells might reflect an erroneous extension of their normal embryonic functions, leading to inappropriate upregulation of cyclin D1, thereby effecting cell cycle regulation and promoting or maintaining an undifferentiated state in breast cells. Overexpression of *cyclin D1* is a common occurrence in mammary carcinogenesis, suggesting a possible link between *MSX* genes and cyclin D1 in breast cancer. Although *MSX* genes might be deregulated in several types of epithelial cancer, their causal effects for inducing carcinogenesis might be restricted to tumor types where their normal downstream target genes, such as *cyclin D1*, have a significant impact.

HSIX1 is the human homologue of the *Drosophila Six* gene that was originally isolated on the basis of its enriched expression during S phase. While not significantly affecting proliferation, ectopic expression of *HSIX1* in cell cultures abrogates the DNA-damage-induced G2 cell-cycle checkpoint. Of particular interest, *HSIX1* is upregulated in primary and metastatic breast cancer, as well as in other tumor types, raising the possibility that its activities in cell culture reflect a relevant role in tumorigenesis *in vivo* [61]. These findings, linking HSIX1 to the regulation of the G2 checkpoint, raise the possibility that deregulation of homeobox genes in cancer might have adverse effects on

the maintenance of genomic stability by allowing cells to proceed inappropriately past the DNA-damage-induced checkpoint.

1.5.3 Loss of homeobox gene expression in carcinoma

There are clearly examples that support the notion where mere 'gains' of expression of certain homeobox genes in cancer cells are oncogenic and that the functional outcome for oncogenesis might vary depending on the cell type in which the particular homeobox gene is expressed. However, there are a few examples where the loss of function of a homeobox gene has been implicated in tumorigenesis. Most notably, methylation of the *HOXA5* promoter, leading to loss of HOXA5 protein expression was correlated with loss of *p53* expression and the subsequent development of breast carcinoma, suggesting that epigenetic silencing of *HOXA5* facilitates breat tumorigenesis [56].

The homeobox gene *CDX2* is an example of homeobox expression in fully differentiated tissue where its expression is involved in maintaining the differentiated state within that tissue. Expression of *Cdx2* first occurs in the developing gut and continues into the adult gut while its expression is lost in colorectal tumors and colon carcinoma cell lines [62]. Re-expression of *Cdx 2* inhibits cell growth and tumorigenicity in established colorectal cell lines while it promotes differentiation of non-tumorigenic epithelial cell lines [63]. Moreover, mice heterozygous for a mutant allele of *Cdx2* frequently develop adenomatous intestinal polyps [64]. Additional genetic events must be required for tumorigenesis since these pre-malignant lesions do not typically progress to the development of overt colon carcinomas. Similar to clinical colon cancer samples, where expression of *CDX2* is frequently lost without a corresponding deleterious mutation, the wildtype *Cdx2* allele remains intact in the precursor polyps from *Cdx2* heterozygous mutant mice despite the loss of Cdx2 protein expression [64]. Thus, epigenetic inactivation of *CDX2*, leading to a loss of *CDX2* expression, could represent an event that predisposes to, but is not sufficient for, the development of overt colorectal cancer.

Analogous to *CDX2*, the homeobox gene *NKX3.1* is expressed during the formation of the prostate and is required for the terminal differentiation of prostate epithelium in the

adult while its loss of function predisposes to cancer in a tissue-specific manner [65]. Overexpression of *Nkx3.1* in prostate carcinoma cells in culture and in nude mice inhibits proliferation and tumorigenicity [66]. There is increased proliferative activity in the prostate epithelium of homozygous and heterozygous *Nkx3.1* mutant mice which subsequently develop prostatic intraepithelial neoplasia (PIN), pre-cancerous lesions of the prostate epithelium that are presumed precursors of prostate cancer [67]. The loss of NKX3.1 function predisposes to, but is not sufficient for, prostate carcinoma [68]. Similar to *CDX2*, the loss of *NKX3.1* expression, due to epigenetic inactivation of one or more of the *NKX3.1* alleles, is involved in human PIN as well as prostate carcinoma [68].

1.5.4 Homeobox genes promote tumorigenesis

Homeobox genes are global regulators of growth and differentiation, with specific members of different classes of homeobox gene families acting at precise developmental stages and in particular tissue types. In normal tissues, the combination of these activities provides the ultimate balance between proliferation and differentiation. Growing evidence supports the broad generalization that the gain and loss of homeobox genes promotes tumorigenesis as a consequence of their inappropriate effects on growth and differentiation, perhaps through an extension of their normal function. Homeobox genes that are negative regulators of differentiation exhibit oncogenic activities when misexpressed while those homeobox genes that are positive regulators of differentiation promote transformation following their loss of function.

The deregulation of homeobox genes appears to promote rather than be sufficient for tumorigenesis, and this deregulation occurs in specific tissues rather than displaying a broad specificity in many tumor types. With the exception of translocations that involve homeobox genes in leukemia [7], deregulation of homeobox genes in solid tumors might not involve mutations or altered functions. With the above characteristics, homeobox genes may be generally defined as positive or negative tumor modulators, rather than as classical oncogenes or tumor suppressors. For many documented cases of perturbations in homeobox expression, a precise or causal relationship between the individual homeobox gene and the tissue-specific cancer phenotype to which it contributes has not

been established. The notion that tissue specificity might reflect the differential activities of normal downstream target genes, as described for *cyclin D1* as a downstream target of *MSX* genes, is starting to emerge. Future studies that will help define the selective functional outcomes of particular homeobox genes and their associated tumor types will no doubt aid in understanding different cancer phenotypes.

1.6 Breast Cancer

Breast cancer is one of the most frequent human malignancies in the Western world. The American Cancer Society estimates that close to 212,930 American women will be diagnosed with breast cancer and that 40,870 women will die from breast cancer this year alone (*www.cancer.org*). Breast cancer is the leading cause of death among women between 40 and 55 years of age and is the second overall cause of death among women from cancer, exceeded only by lung cancer. Fortunately, the mortality rate from breast cancer is slowly decreasing with an increased emphasis on early detection and more effective treatments. The pathogenesis of this disease is thought to involve multiple genetic and epigenetic events. In spite of recent advances in the assessment of breast cancer risk, through the identification of crucial susceptibility genes (*BRCA1/2*, *PTEN*, *P53*), these account for less than 5% of all breast cancer cases and may not be associated with the more commonly occurring sporadic breast cancers [69]. Sporadic nonhereditary breast cancer is recognized as the most common form of this malignancy.

Like most other cancers, it is thought that breast cancer originates in one cell which acquires numerous events and becomes malignant. Later additional events lead to the development of different clones and with different characteristics. Few genes have been found to be mutated in breast tumors but several chromosome arms with yet identified genes have been found to be involved. These aberrations include DNA amplifications and deletions. Genetic abnormalities in breast cancer include amplification of oncogenes, mutation of tumor suppressor genes and loss of heterozygosity (LOH) at chromosomes 1, 3p, 6q, 7q, 8p, 9p, 10q, 11, 13q, 16q, 17, 18q, 22q and X [70]. Some of these genetic abnormalities exist in premalignant breast tissue and may have a causative

role in breast tumor pathogenesis. Unike colon cancer, the sequential steps of molecular lesions during breast cancer progression are poorly understood.

Only a few critical oncogenes have been identified in sporadic breast cancer and the discovery of bona fide primary genetic lesions underlying sporadic breast cancer development remains a major challenge. This is due, at least in part, to the marked cytogenetic complexity seen in advanced breast cancer, precluding investigators from readily identifying primary causative genetic events in breast cell transformation [71]. The *c-MYC* gene is amplified and/or overexpressed in a high proportion of human breast cancer, although the frequency of these alterations varies greatly [72]. *ErbB2* is also amplified and subsequently overexpressed in 20-30% of human breast cancers, and overexpression of *ErbB2* is correlated with a poor clinical prognosis of both node-positive and node-negative tumors [73]. The *Cyclin D1* gene is amplified in 15-20% of human breast cancers [74]. As the oncogenes located at amplified chromosomal regions are rarely amplified in benign breast disease [75], they may represent later events in the multistep progression associated with the development of breast cancer.

One could argue that the only known and well validated human breast cancer oncogene is *ErbB2* since it is the only gene that has been shown to strongly correlate with human breast cancer progression in studies of primary clinical samples, to be transforming in human mammary epithelial cells *in vitro*, to cause mammary tumor development in transgenic mice [76] and most importantly, for which a targeted therapeutic has been developed that has clear efficacy in the clinic [77]. Thus, *ErbB2* is the strongest candidate for a bona fide human breast cancer oncogene and is based both on correlative data and convincing mechanistic data.

Evidence is accumulating that suggests the breast and ovarian cancer susceptibility genes *BRCA1* and *BRCA2* seem to be involved in sporadic cancer as well. Recent evidence suggests that *BRCA1* methylation contributes to a small subset of sporadic breast cancer and the resulting molecular and clinical phenotype is similar to that of hereditary *BRCA1*-associated breast cancers. A model has now been proposed where

BRCA1 promoter methylation may serve as a 'first hit' in a fraction of sporadic breast carcinogenesis, much like an inherited germ line mutation, which promotes tumor progression [78]. The BRCA2-interacting protein EMSY, might now also provide the link between the *BRCA2* pathway and sporadic cancer. EMSY and BRCA2 have overlapping functions and *EMSY* was found to be amplified in 5/28 (18%) breast cancer cell lines and in 70/551 (13%) poor-prognosis node-negative breast cancer cases [79]. EMSY was shown to inhibit the transcriptional activation function of BRCA2, so it is conceivably possible that *BRCA2* deletion and *EMSY* amplification have similar effects [79].

Breast cancer research in the past has focused on identifying mutations in breast tumors and then determining their role in mammary carcinogenesis. These efforts have clearly shown that breast cancer is a genetic disease where changes in the genome can lead to putative causative events in the development and progression of mammary carcinogenesis. Recent studies like those above with BRCA1 and BRCA2 in sporadic cases also clearly indicate that the phenotype of most breast cancers, if not all cancers, is probably due to an amalgamation of some mutated genes and some genes that are functionally modified by epigenetic changes such as hypermethylation of DNA and hypoacetylation of histones [80]. Inactivation or silencing of the *E-cadherin* gene is associated with increased tumor invasiveness and is often found in late stage carcinomas [81]. Hypermethylation of the *E-cadherin* promoter is the principal mechanism of its loss in many types of cancer, and its occurence is thought to reflect the stage of tumor progression [81].

The abundance of important genes identified as being epigenetically modified in cancer has led to the expansion of Knudson's 'two-hit' hypothesis for loss of tumor suppressor gene function in tumor development to include epigenetic mechanisms as bona fide 'hits'. Whether a mutation or epigenetic event occurs first is unknown. It is known that the development of sporadic breast cancer is driven by phenotypic changes due to both genetic and epigenetic events. One type of event may suffice in initiating the disease but however initiated, the heterogeneity of sporadic breast cancer suggests that the etiology of each tumor may be specific to the individual.

Epigenetic dysregulation can also lead to aberrant expression or overexpression. For example, the lysosomal protease cathepsin D is a predictor of poor outcome in breast cancer patients and the genes 5' regulatory region displays differences in chromatin structure in hormone-dependent vs. hormone-independent breast cancer cell lines [82]. Differences in *cathepsin D* expression in breast cancer may conceivably involve changes in chromatin architecture, which could render its expression constitutive. The gene for the intermediate filament protein vimentin is transcriptionally ectopically activated during tumor progression and also displays differences in chromatin architecture in vimentin-positive vs. vimentin-negative breast cancer cells [83].

1.6.1 Breast Cancer Stem Cells

Stem cells have a large replicative potential and a long life span, which makes them excellent candidates for the cells of origin of breast tumors. Not only will MSCs persist in the body for long enough to accumulate the many mutations that are required to change a normal cell into one with neoplastic potential, they also have the proliferative capacity to easily generate a tumor mass. It is generally accepted that most tumors are clonal in origin and represent the progeny of a single cell but it was unclear which cells in the tumor had the capacity to maintain and regenerate tumor growth. Two theories were proposed to explain why only certain cells within a tumor could initiate tumor growth. The stochastic model predicted that the tumor is relatively homogeneous and the genetic alterations that underlie tumorigenesis were operative and shared among all cells within the tumor and although each cell is potentially tumor-initiating, the necessary entry into the cell cycle is governed by low probability stochastic events [84]. The hierarchy model predicts functional heterogeneity among the cells that make up the tumor and that there are distinct and rare tumor initiating cells within the tumor.

A major step towards identifying these latter breast cancer-initiating cells was provided by elegant studies that support the theory that breast cancer is functionally heterogeneous and that only a rare breast cancer-initiating cell type is capable of establishing human breast cancer after being transplanted. These studies showed that a

minor breast tumor stem cell like subpopulation was found to have a 10-50-fold increase in the ability to form tumors in xenografts compared with the bulk of breast tumor cells [85]. This subpopulation represents 2% of the unfractionated breast cancer cells derived from human breast tumor samples and lacks differentiated breast epithelial cell lineage markers (Lineage-) and was isolated using specific cell surface markers; CD44 and CD24 adhesion molecules and the epithelial specific antigen ESA. The expression of these markers is heterogeneous in breast cancer tissue but the breast cancer-initiating tumor cells were found to reside in a distinct $Lin^- ESA^+ CD44^+ CD24^{-/low}$ population which can be isolated by cell fractionation [85].

Evidence supporting the existence of a breast tumor stem cell population has further fueled the search for the putative but elusive adult mammary epithelial stem cell in the normal breast. Identifying the existence of these tumor-initiating cells within breast tumors also has enormous implications for how we approach tumor therapy as most traditional cancer treatments target rapidly proliferating cells. Although this might eliminate the non-tumor-initiating cancer cells that make up the bulk of a tumor and therefore lead to remission, the relatively quiescent tumor-initiating stem cells might be resistant to traditional chemotherapy, allowing the disease to relapse if they are not eliminated.

Whether breast tumor stem cells originate from normal adult MSCs or represent a transit or progenitor cell population in the normal breast that has gained stem cell like properties through the possible induction of cell type de-differentiation remains to be answered. If normal MSCs are the primary target in the process of tumorigenesis, the resulting tumors could be poorly differentiated and highly aggressive while tumors arising from transit or progenitor cells may be relatively well differentiated and noninvasive. Much of the proliferative capacity of normal stem cells could very well reside in a progenitor cell population however, since these daughter cells could conceivably maintain replicative potential and long life by means of a slow rate of division.

1.6.2 Breast cancer stem cells as therapeutic targets

The advantage of targeting breast cancer stem cells is that tumors could be treated on the basis of a characteristic property shared by all breast tumor stem cells, one that is absent from normal mammary stem cells. Instead of dissecting each and every genetic defect and customizing a treatment for each individual and tumor type, this type of treatment could enable the elimination of the cells of origin of the tumor, and make promising therapeutic targets since these cells are those that likely remain after most regimens and lead to eventual relapses in some patients.

If breast tumor stem cells are derived from progenitor or differentiated cells instead of normal adult MSCs, they are likely to still have phenotypic characteristics of normal MSCs. Reversion of progenitor or differentiated cells to a more stem cell like state is thought to be one possible scenario leading to the neoplastic process [86]. In this scenario, the phenotypic characteristics of normal MSCs might still be effectively employed to target breast tumor cells [86]. Future and ongoing studies are aimed at identifying distinct molecules that are involved in the regulation of MSCs and patterns of gene-expression sets that are limited to and define MSCs, as well as genes that are potentially common to all stem cells.

1.7 Objectives

The objective of the research summarized in the present PhD thesis was to identify a gene involved in mammary tumorigenesis and to investigate whether this gene was also involved in normal mammary gland development. The work presented herein represents a candidate gene approach, which focused on the *Engrailed* family of homeobox genes because of the following motivations:

(i) *Engrailed-1 (En-1)* and *Engrailed-2 (En-2)* genes are downstream targets of *Wnt-1* signaling during mouse embryogenesis that have important roles in development. Ectopic expression of *Wnt-1* under the control of the mouse mammary tumor virus (MMTV) promoter leads to extensive mammary hyperplasia and the subsequent

generation of the majority of MMTV-induced adenocarcinomas [87]. Although expression of *WNT1* itself has not been reported in normal or neoplastic human breast tissue, other *WNT* family members are expressed in a proportion of both normal human breast tissue and breast cancer samples [88]. *Wnt-1* encodes a secreted growth factor that initiates a signaling cascade, which results in transcriptional activation mediated by β-catenin/Tcf complexes [40]. β-catenin/Tcf-mediated transcription has been implicated in several human cancers, with some targets relevant to breast carcinogenesis identified such as c-*MYC* and *cyclin D1* [89]. Also at the time this doctoral reseach was initiated, the observation that *Wnt-1* is a mouse mammary oncogene and that the downstream mediator β-catenin is often stabilized in certain human malignancies fueled the ongoing search for additional targets of this pathway, such as the *Engrailed* genes, that might also be implicated in human breast cancer;

(ii) The mammary gland is established during embryogenesis but the majority of mammary gland growth and development occur in the adult female. *En* genes are homeodomain-containing transcriptional regulators that have been studied extensively in embryonic development, but they are also part of a growing number of transcription factors which have evolved to carry out additional functions in the adult mammal. At the time this thesis was initiated, reports were beginning to provide evidence that *En-1* had indeed acquired a further role in postnatal limb development and thus it is a possibility that these genes have evolved to carry out additional postnatal roles, such as in postnatal mammary gland development;

(iii) Studies were being published at the time that reported, similar to several Hox proteins, that Pbx is a co-operative DNA-binding partner of En-2 [5]. In addition, several homeodomain proteins have been implicated as causative oncogenes in cancer and studies in the lab were showing that Pbx was enhancing the transforming ability of Hox genes when overexpressed *in vitro*.

The research presented in this PhD thesis describes efforts towards investigating whether *En-1* or *En-2* were expressed in normal breast tissue, immortalized mammary epithelial

cell lines and primary breast tumors and whether they might be involved in regulating critical aspects of mammary gland development and/or mammary tumorigenesis. Chapter 2 summarizes the identification of a putative role for *En-1* in postnatal mouse mammary gland development while Chapter 3 documents studies that suggest *EN2* is a candidate oncogene in human breast cancer. Chapter 4 summarizes the identification of putative *EN2* transcriptional targets by microarray analysis. Chapter 5 concludes with a discussion of the relevance of the findings presented in this thesis and addresses potential future studies that could provide further insight into the roles of *En-1* and *EN2* in the regulation of normal mammary gland development and breast cancer, respectively.

References

1. Akin,Z.N. & Nazarali,A.J. Hox genes and their candidate downstream targets in the developing central nervous system. *Cell Mol.Neurobiol.* **25**, 697-741 (2005).

2. Messina,D.N., Glasscock,J., Gish,W. & Lovett,M. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res.* **14**, 2041-2047 (2004).

3. Lewis,M.T. Homeobox genes in mammary gland development and neoplasia. *Breast Cancer Res.* **2**, 158-169 (2000).

4. Kissinger,C.R. et al. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. *Cell.* **63**, 579-590 (1990).

5. Peltenburg,L.T. & Murre,C. Specific residues in the Pbx homeodomain differentially modulate the DNA-binding activity of Hox and Engrailed proteins. *Development.* **124**, 1089-1098 (1997).

6. Williams,T.M., Williams,M.E. & Innis,J.W. Range of HOX/TALE superclass associations and protein domain requirements for HOXA13:MEIS interaction. *Dev.Biol.* **277**, 457-471 (2005).

7. Abate-Shen,C. Deregulated homeobox gene expression in cancer: cause or consequence? *Nat.Rev.Cancer.* **2**, 777-785 (2002).

8. Bendall,A.J. & Abate-Shen,C. Roles for Msx and Dlx homeoproteins in vertebrate development. *Gene.* **247**, 17-31 (2000).

9. Harvey,R.P. NK-2 homeobox genes and heart development. *Dev.Biol.* **178**, 203-216 (1996).

10. Prakash,N. & Wurst,W. Specification of midbrain territory. *Cell Tissue Res.* **318**, 5-14 (2004).

11. Chapouton,P. & Bally-Cuif,L. Neurogenesis. *Methods Cell Biol.* **76**, 163-206 (2004).

12. Marie,B. & Bacon,J.P. Two engrailed-related genes in the cockroach: cloning, phylogenetic analysis, expression and isolation of splice variants. *Dev.Genes Evol.* **210**, 436-448 (2000).

13. Simon,H.H., Thuret,S. & Alberi,L. Midbrain dopaminergic neurons: control of their cell fate by the engrailed transcription factors. *Cell Tissue Res.* **318**, 53-61 (2004).

14. Kornberg,T. Engrailed: a gene controlling compartment and segment formation in Drosophila. *Proc.Natl.Acad.Sci.U.S.A.* **78**, 1095-1099 (1981).

15. Joyner,A.L. & Martin,G.R. En-1 and En-2, two mouse genes with sequence homology to the Drosophila engrailed gene: expression during embryogenesis. *Genes Dev.* **1**, 29-38 (1987).

16. Force,A. et al. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics.* **151**, 1531-1545 (1999).

17. Hanks,M. et al. Rescue of the En-1 mutant phenotype by replacement of En-1 with En-2. *Science.* **269**, 679-682 (1995).

18. Logan,C. et al. Cloning and sequence comparison of the mouse, human, and chicken engrailed genes reveal potential functional domains and regulatory regions. *Dev.Genet.* **13**, 345-358 (1992).

19. Tolkunova,E.N. et al. Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. *Mol.Cell Biol.* **18**, 2804-2814 (1998).

20. Peltenburg,L.T. & Murre,C. Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J.* **15**, 3385-3393 (1996).

21. Gemel,J., Jacobsen,C. & MacArthur,C.A. Fibroblast growth factor-8 expression is regulated by intronic engrailed and Pbx1-binding sites. *J.Biol.Chem.* **274**, 6020-6026 (1999).

22. Kobayashi,M. et al. Engrailed cooperates with extradenticle and homothorax to repress target genes in Drosophila. *Development.* **130**, 741-751 (2003).

23. Serrano,N. & Maschat,F. Molecular mechanism of polyhomeotic activation by Engrailed. *EMBO J.* **17**, 3704-3713 (1998).

24. Cosgaya,J.M., Aranda,A., Cruces,J. & Martin-Blanco,E. Neuronal differentiation of PC12 cells induced by engrailed homeodomain is DNA-binding specific and independent of MAP kinases. *J.Cell Sci.* **111 (Pt 16)**, 2377-2384 (1998).

25. Maizel,A. et al. Engrailed homeoprotein secretion is a regulated process. *Development.* **129**, 3545-3553 (2002).

26. Joliot,A. et al. Association of Engrailed homeoproteins with vesicles presenting caveolae-like properties. *Development.* **124**, 1865-1875 (1997).

27. Joliot,A. et al. Identification of a signal sequence necessary for the unconventional secretion of Engrailed homeoprotein. *Curr.Biol.* **8**, 856-863 (1998).

28. Maizel,A., Bensaude,O., Prochiantz,A. & Joliot,A. A short region of its homeodomain is necessary for engrailed nuclear export and secretion. *Development.* **126**, 3183-3190 (1999).

29. McMahon,A.P., Joyner,A.L., Bradley,A. & McMahon,J.A. The midbrain-hindbrain phenotype of Wnt-1-/Wnt-1- mice results from stepwise deletion of engrailed-expressing cells by 9.5 days postcoitum. *Cell.* **69**, 581-595 (1992).

30. Davis,C.A., Holmyard,D.P., Millen,K.J. & Joyner,A.L. Examining pattern formation in mouse, chicken and frog embryos with an En-specific antiserum. *Development.* **111**, 287-298 (1991).

31. Wurst,W., Auerbach,A.B. & Joyner,A.L. Multiple developmental defects in Engrailed-1 mutant mice: an early mid-hindbrain deletion and patterning defects in forelimbs and sternum. *Development.* **120**, 2065-2075 (1994).

32. Loomis,C.A. et al. The mouse Engrailed-1 gene and ventral limb patterning. *Nature.* **382**, 360-363 (1996).

33. Wurst,W., Auerbach,A.B. & Joyner,A.L. Multiple developmental defects in Engrailed-1 mutant mice: an early mid-hindbrain deletion and patterning defects in forelimbs and sternum. *Development.* **120**, 2065-2075 (1994).

34. Millen,K.J., Wurst,W., Herrup,K. & Joyner,A.L. Abnormal embryonic cerebellar development and patterning of postnatal foliation in two mouse Engrailed-2 mutants. *Development.* **120**, 695-706 (1994).

35. Joyner,A.L., Skarnes,W.C. & Rossant,J. Production of a mutation in mouse En-2 gene by homologous recombination in embryonic stem cells. *Nature.* **338**, 153-156 (1989).

36. Hanks,M.C. et al. Drosophila engrailed can substitute for mouse Engrailed1 function in mid-hindbrain, but not limb development. *Development.* **125**, 4521-4530 (1998).

37. Danielian,P.S. & McMahon,A.P. Engrailed-1 as a target of the Wnt-1 signalling pathway in vertebrate midbrain development. *Nature.* **383**, 332-334 (1996).

38. Klingensmith,J., Nusse,R. & Perrimon,N. The Drosophila segment polarity gene dishevelled encodes a novel protein required for response to the wingless signal. *Genes Dev.* **8**, 118-130 (1994).

39. Liu,A. & Joyner,A.L. Early anterior/posterior patterning of the midbrain and cerebellum. *Annu.Rev.Neurosci.* **24**, 869-896 (2001).

40. Cadigan,K.M. & Nusse,R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11**, 3286-3305 (1997).

41. Hennighausen,L. & Robinson,G.W. Information networks in the mammary gland. *Nat.Rev.Mol.Cell Biol.* **6**, 715-725 (2005).

42. Hennighausen,L. & Robinson,G.W. Think globally, act locally: the making of a mouse mammary gland. *Genes Dev.* **12**, 449-455 (1998).

43. Humphreys,R.C. et al. Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis. *Development.* **122**, 4013-4022 (1996).

44. DeOme,K.B. et al. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *J.Natl.Cancer Inst.* **78**, 515-525 (1959).

45. Daniel,C.W. et al. The in vivo life span of normal and preneoplastic mouse mammary glands: a serial transplantation study. *Proc.Natl.Acad.Sci.U.S.A.* **61**, 53-60 (1968).

46. Chepko,G. & Smith,G.H. Three division-competent, structurally-distinct cell populations contribute to murine mammary epithelial renewal. *Tissue Cell.* **29**, 239-253 (1997).

47. Stingl,J. et al. Purification and unique properties of mammary epithelial stem cells. *Nature.* **439**, 993-997 (2006).

48. Kordon,E.C. & Smith,G.H. An entire functional mammary gland may comprise the progeny from a single cell. *Development.* **125**, 1921-1930 (1998).

49. Smith,G.H. Experimental mammary epithelial morphogenesis in an in vivo model: evidence for distinct cellular progenitors of the ductal and lobular phenotype. *Breast Cancer Res.Treat.* **39**, 21-31 (1996).

50. Dontu,G. et al. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* **17**, 1253-1270 (2003).

51. Duboule,D. No milk today (my Hox have gone away). *Proc.Natl.Acad.Sci.U.S.A.* **96**, 322-323 (1999).

52. Satokata,I., Benson,G. & Maas,R. Sexually dimorphic sterility phenotypes in Hoxa10-deficient mice. *Nature.* **374**, 460-463 (1995).

53. Friedmann,Y., Daniel,C.A., Strickland,P. & Daniel,C.W. Hox genes in normal and neoplastic mouse mammary gland. *Cancer Res.* **54**, 5981-5985 (1994).

54. Chen,F. & Capecchi,M.R. Paralogous mouse Hox genes, Hoxa9, Hoxb9, and Hoxd9, function together to control development of the mammary gland in response to pregnancy. *Proc.Natl.Acad.Sci.U.S.A.* **96**, 541-546 (1999).

55. Phippard,D.J. et al. Regulation of Msx-1, Msx-2, Bmp-2 and Bmp-4 during foetal and postnatal mammary gland development. *Development*. **122**, 2729-2737 (1996).

56. Raman,V. et al. Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature*. **405**, 974-978 (2000).

57. Kozmik,Z. et al. Deregulated expression of PAX5 in medulloblastoma. *Proc.Natl.Acad.Sci.U.S.A.* **92**, 5709-5713 (1995).

58. Bowen,C. et al. Loss of NKX3.1 expression in human prostate cancers correlates with tumor progression. *Cancer Res.* **60**, 6111-6115 (2000).

59. Krosl,J. et al. Cellular proliferation and transformation induced by HOXB4 and HOXB3 proteins involves cooperation with PBX1. *Oncogene*. **16**, 3403-3412 (1998).

60. Hu,G. et al. Msx homeobox genes inhibit differentiation through upregulation of cyclin D1. *Development*. **128**, 2373-2384 (2001).

61. Ford,H.L. et al. Abrogation of the G2 cell cycle checkpoint associated with overexpression of HSIX1: a possible mechanism of breast carcinogenesis. *Proc.Natl.Acad.Sci.U.S.A.* **95**, 12608-12613 (1998).

62. Hinoi,T. et al. Loss of CDX2 expression and microsatellite instability are prominent features of large cell minimally differentiated carcinomas of the colon. *Am.J.Pathol.* **159**, 2239-2248 (2001).

63. Mallo,G.V. et al. Expression of the Cdx1 and Cdx2 homeotic genes leads to reduced malignancy in colon cancer-derived cells. *J.Biol.Chem.* **273**, 14030-14036 (1998).

64. Chawengsaksophak,K. et al. Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature*. **386**, 84-87 (1997).

65. Bhatia-Gaur,R. et al. Roles for Nkx3.1 in prostate development and cancer. *Genes Dev.* **13**, 966-977 (1999).

66. Kim,M.J. et al. Nkx3.1 mutant mice recapitulate early stages of prostate carcinogenesis. *Cancer Res.* **62**, 2999-3004 (2002).

67. Abdulkadir,S.A. et al. Conditional loss of Nkx3.1 in adult mice induces prostatic intraepithelial neoplasia. *Mol.Cell Biol.* **22**, 1495-1503 (2002).

68. Kim,M.J. et al. Cooperativity of Nkx3.1 and Pten loss of function in a mouse model of prostate carcinogenesis. *Proc.Natl.Acad.Sci.U.S.A.* **99**, 2884-2889 (2002).

69. Baselga,J. & Norton,L. Focus on breast cancer. *Cancer Cell.* **1**, 319-322 (2002).

70. Ingvarsson,S. Molecular genetics of breast cancer progression. *Semin.Cancer Biol.* **9**, 277-288 (1999).

71. Tognon,C. et al. Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell.* **2**, 367-376 (2002).

72. Liao,D.J. & Dickson,R.B. c-Myc in breast cancer. *Endocr.Relat Cancer.* **7**, 143-164 (2000).

73. Hutchinson,J.N. & Muller,W.J. Transgenic mouse models of human breast cancer. *Oncogene.* **19**, 6130-6137 (2000).

74. Bieche,I. et al. Quantitation of MYC gene expression in sporadic breast tumors with a real-time reverse transcription-PCR assay. *Cancer Research.* **59**, 2759-2765 (1999).

75. Lizard-Nacol,S. et al. Benign breast disease: absence of genetic alterations at several loci implicated in breast cancer malignancy. *Cancer Res.* **55**, 4416-4419 (1995).

76. Muller,W.J. et al. Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated c-neu oncogene. *Cell.* **54**, 105-115 (1988).

77. Ethier,S.P. Identifying and validating causal genetic alterations in human breast cancer. *Breast Cancer Res.Treat.* **78**, 285-287 (2003).

78. Wei,M. et al. BRCA1 promoter methylation in sporadic breast cancer is associated with reduced BRCA1 copy number and chromosome 17 aneusomy. *Cancer Res.* **65**, 10692-10699 (2005).

79. Hughes-Davies,L. et al. EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell.* **115**, 523-535 (2003).

80. Mielnicki,L.M., Asch,H.L. & Asch,B.B. Genes, chromatin, and breast cancer: an epigenetic tale. *J.Mammary.Gland.Biol.Neoplasia.* **6**, 169-182 (2001).

81. Nass,S.J. et al. Aberrant methylation of the estrogen receptor and E-cadherin 5' CpG islands increases with malignant progression in human breast cancer. *Cancer Res.* **60**, 4346-4348 (2000).

82. Giamarchi,C. et al. Chromatin structure of the regulatory regions of pS2 and cathepsin D genes in hormone-dependent and -independent breast cancer cell lines. *Oncogene.* **18**, 533-541 (1999).

83. Sommers,C.L. et al. Regulation of vimentin gene transcription in human breast cancer cell lines. *Cell Growth Differ.* **5**, 839-846 (1994).

84. Dick,J.E. Breast cancer stem cells revealed. *Proc.Natl.Acad.Sci.U.S.A.* **100**, 3547-3549 (2003).

85. Al Hajj,M. et al. From the Cover: Prospective identification of tumorigenic breast cancer cells. *Proc.Natl.Acad.Sci.U.S.A.* **100**, 3983-3988 (2003).

86. Smalley,M. & Ashworth,A. Stem cells and breast cancer: A field in transit. *Nat.Rev.Cancer.* **3**, 832-844 (2003).

87. Nusse,R. & Varmus,H.E. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell.* **31**, 99-109 (1982).

88. Huguet,E.L. et al. Differential expression of human Wnt genes 2, 3, 4, and 7B in human breast cell lines and normal and disease states of human breast tissue. *Cancer Res.* **54**, 2615-2621 (1994).

89. He,T.C. et al. Identification of c-MYC as a target of the APC pathway. *Science.* **281**, 1509-1512 (1998).

CHAPITRE 2

ARTICLE

En-1 Deficiency Leads to Abnormal Ductal Development In the Mouse Mammary Gland

Nicole L. Martin[1] and Guy Sauvageau[1,2,3] *

Accepted pending minor modifications in *Developmental Dynamics* (2006)

1Laboratory of Molecular Genetics of Hematopoietic Stem Cells, Institut de Recherches en Immunologie et Cancérologie, C.P. 6128, succursale Centre-Ville, Montréal, Québec, Canada, H3C 3J7; 2Department of Medicine and 3Department of Medicine and Division of Hematology and Leukemia Cell Bank of Quebec, Maisonneuve-Rosemont Hospital, Montréal, Québec, Canada, H1T 2M4.

Chapter 2 describes a study which documents the unique spatio-temporal pattern of *En-1* expression in the mammary epithelium of the prepubertal and early pubertal mouse mammary gland and suggests a potential role for *En-1* in the initial growth and morphogenesis of the epithelial ductal system during the onset of puberty. Nicole Martin generated all the results presented in this chapter and wrote the paper under the supervision of Dr. Guy Sauvageau.

Abstract

We have defined the developmentally regulated expression profile of *Engrailed-1 (En-1)* and the putative cell types which express this gene in the mammary epithelium of the prepubertal and early pubertal mouse mammary gland. Using *En-1* mutant mice, we provide evidence that loss of En-1 function in the female mammary gland results in severely impaired ductal growth. Pubertal *En-1* null mammary glands revealed a primitive ductal rudiment devoid of terminal end buds, reminiscent of a prepubertal mammary gland, while a fully developed ductal system was seen in *En-1* heterozygous and wildtype siblings. Moreover, *En-1* null mammary epithelium transplanted into surgically cleared fat pads of wildtype syngeneic hosts displayed limited ductal outgrowth with reduced side branching. Together, these studies demonstrate a unique spatio-temporal pattern of *En-1* expression in mammary tissue and suggest a potential contribution for this gene in the initial growth and morphogenesis of the epithelial ductal system during the onset of puberty.

Introduction

Although the mammary gland is established during fetal development, most of its expansion and development occurs postnatally [1]. At birth, the mammary epithelial tree emanates from a central primary duct, which is connected to the nipple. Functional development of the postnatal mammary gland proceeds in distinct stages. At roughly 3-4 weeks of age, ovarian hormones stimulate the accelerated and invasive ductal extension and branching of the mammary tree. The onset of puberty also coincides with the appearance of large club-shaped terminal end buds (TEBs), highly proliferative structures found at the tips of the growing ducts. It is the bifurcation of these TEBs through the stromal fat pad that creates the branched pattern of the ductal system found within the mature mammary gland. The TEB is a specialized structure composed of two distinct and relatively undifferentiated cell types. Cap cells make up the outermost layer of the TEB and interact with the surrounding stroma as the subtending duct is formed while body cells (about 6-10 layers), fill the interior of the TEB. Cap cells are the precursors of myoepithelial cells, which form an outer sleeve around the primary ducts and become discontinuous around secondary and tertiary ducts. The inner body cells give rise to luminal epithelial cells. A single layer of luminal epithelial cells will line the newly forming ducts, where the ducts will eventually serve as channels for milk transport during lactation. Temporally, the ductal system reaches the central lymph node around 4 ½ weeks and continues growing until the ducts have reached the limits of the mammary fat pad around 8 weeks of age. At this point the mammary gland reaches maturity, ductal elongation ceases and TEBs regress to leave a branched system of differentiated ducts. Lobulo-alveolar development and proliferation occur during pregnancy, and functional differentiation of the secretory epithelium coincides with parturition and lactation. After weaning, the entire mammary alveolar compartment is remodeled to eventually resemble that of a mature virgin gland. With each subsequent pregnancy, a new cycle of lobulo-alveolar development occurs.

An essential feature of the mouse mammary gland is the regenerative capacity of its epithelium. Portions of the epithelial tree can be transplanted into the mammary fat pad of a syngeneic female whose endogenous mammary epithelium has been removed, and

reproduce a complete and functional mammary gland [2]. The ability to delete genes from the mouse genome, in conjunction with tissue transplants to evaluate their physiologic role, has led to the identification of several genes involved in mammary gland development. One family of genes that has been proposed to be capable of regulating such developmental decisions in the growth and differentiation of this tissue is the homeobox family.

We have recently shown that the homeobox gene *Engrailed-2* (*EN2)* is not expressed in the normal mouse or human mammary tissue, but is aberrantly expressed in a subset of human breast cancer and induces mammary tumors in mice [3]. In addition, RNA interference studies showed the importance of *EN2* expression in maintaining the transformed phenotype of a human breast cancer cell line [3]. Although both *En-2,* and its paralog *En-1,* readily transform mammary epithelial cell lines, *En-1* was not detected in any of the primary breast tumors analyzed (n=82). Both *En-1* and *En-2* are required for midbrain and cerebellum development and in addition, *En-1* plays a crucial role in dorsal/ventral patterning of the limbs and skeleton [4].

The fact that *EN2* has been shown to promote mammary tumorigenesis, combined with evidence indicating that *En-1* has evolved to perform further critical functions in postnatal limb development in addition to its essential role in embryonic development, prompted us to investigate whether *En-1* has acquired an additional role in mammary gland development. The role of En-1 in neural and limb development has been studied extensively, yet the determination of a potential role in postnatal mammary gland development has been hampered by the fact that *En-1* null mutants die shortly after birth with a large mid-hindbrain deletion and skeletal defects of the limbs, 13th rib and sternum [4]. In the present study we employed *En-1*$^{Lki/+}$ (*Lki*, *lacZ* knock-in) transgenic mice in which the bacterial reporter gene *lacZ* is under the control of the *En-1* promoter and endogenous regulatory elements, to demonstrate cell type specific and developmentally regulated expression of *En-1* in mammary epithelial cells during the onset of puberty, corresponding to a phase of overt proliferation and morphogenesis in the mouse mammary gland. The targeting of *lacZ* into the first exon of *En-1* produces a

null allele in *En-1*, and rare surviving mutant *En-1*$^{Lki/Lki}$ females revealed a primitive ductal rudiment devoid of TEBs while a fully developed ductal tree was seen in the heterozygous and wildtype siblings. Moreover, mammary epithelium from *En-1*$^{Lki/Lki}$ females was transplanted into the cleared fat pads of syngeneic 3 week old recipients where the resulting transplants displayed impaired ductal outgrowth and branching morphogenesis. These observations suggest that *En-1* may be involved in the initial growth and expansion of the epithelial ductal system during mammary gland development.

Results

Temporal expression of *En-1* in the mouse mammary gland

An initial investigation of En-1 protein expression in the mouse mammary gland revealed that while En-1 was not detectable in the mammary tissue derived from a lactating mother, it was readily detected in the mammary tissue of both of her 17-day old suckling female pups (lane 6-7 in Fig. 2.1a). This observation was confirmed using RT-PCR analysis, which also detected very low levels of *En-1* expression in the mammary tissue of a 35-day old female, but not in any of the later pubertal timepoints, or at any developmental stage during pregnancy, lactation or involution (data not shown). Mammary tissue isolated from additional prenatal and postnatal developmental stages showed that En-1 was first detected in the mammary tissue of a 15-day old prepubertal female and expression levels steadily increase until puberty commences around day 24-28, and become undetectable again at day 35 during puberty (Fig. 2.1b, c). En-1 protein was not detected in embryonic mammary tissue, mammary tissue containing the entire mammary epithelium from a 17-day old suckling male pup or in mammary tissue isolated during pregnancy or involution (Fig. 2.1b, c). Moreover, En-1 was not detectable in the fatty stroma isolated from a 24-day old early pubertal female gland in which the epithelial tissue was surgically removed, while the highest levels of En-1 protein were observed in the tissue portion which contained the mammary epithelial tree (Fig. 2.1b, c: lane 7 and 4, for epithelial-enriched and -depleted tissue, respectively). Of

interest, this highly regulated expression of *En-1* in the prepubertal and early pubertal mammary gland coincides with the rapid expansion of the epithelial ductal system and the appearance of TEBs.

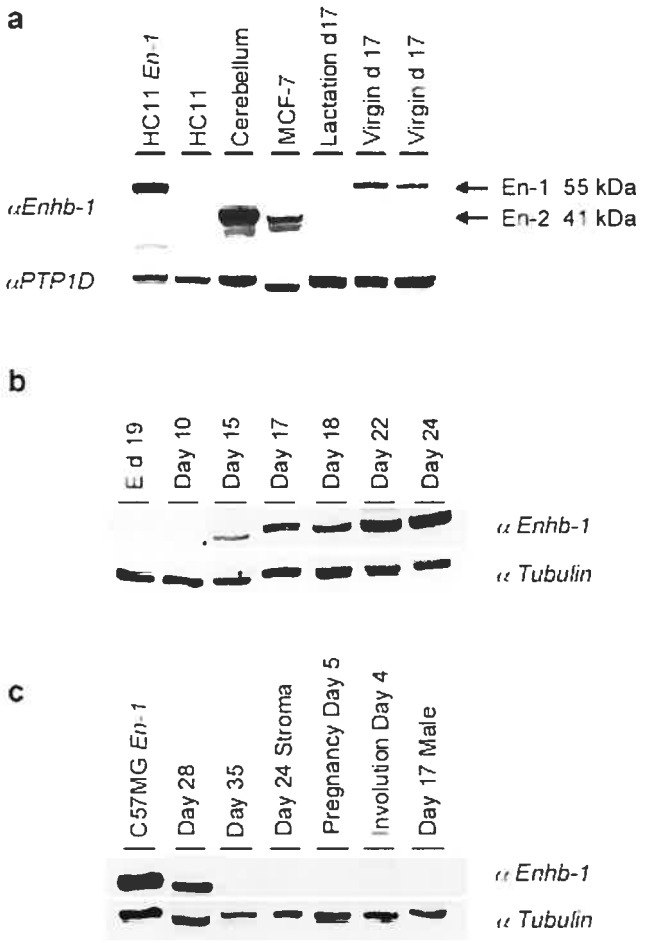

Fig. 2.1 En-1 protein is detected in the prepubertal and early pubertal female mammary gland. Western blot analysis of En-1 protein expression in mammary tissue derived from: (a) a nursing mother at lactation day 17 (Lactation d17), and two of the suckling 17-day old female pups (Virgin d17) and, (b-c) female inguinal mammary glands at the following developmental timepoints; embryonic day 19 (Ed19), 10 to 35-day old prepubertal and pubertal immature virgin glands (Day 10 to Day 35), 5 days of pregnancy (Pregnancy Day 5), and 4 days after weaning (Involution). In addition, protein was extracted from the entire mammary gland isolated from a 17-day old

prepubertal virgin male (Day 17 Male). The portion of the fat pad beyond the lymph node where the epithelial tree has not yet penetrated was surgically removed and was included to represent epithelial-free mammary fatty tissue (Day 24 Stroma). HC11 and C57MG mammary epithelial cells engineered to express *En-1,* the MCF-7 human breast-tumor derived cell line that expresses endogenous *EN2* and adult mouse cerebellum which expresses *En-2* were included as specificity controls while the Pregnant and Involution samples were included as negative controls. α *Enhb-1*; antiserum that recognizes both En-1 and En-2, *αPTP1D*; protein-tyrosine phosphatase 1D antibody, α *Tubulin*; recognizes 55 kDa beta Tubulin.

_En-1_ is expressed in cells lining the mammary epithelial tree during puberty

To localize the spatial expression pattern of *En-1* in the early pubertal mammary gland, the targeted mouse line, $En\text{-}1^{Lki/+}$, in which *En-1* was functionally replaced with the bacterial *lacZ* reporter gene, was employed (Fig. 2.2a, b) [5]. Although the targeting event that brought the integrated *lacZ* sequence under the control of the *En-1* locus also introduced a null mutation in *En-1*, $En\text{-}1^{Lki/+}$ heterozygous mice are viable phenocopies of wildtype littermates and females display normal mammary gland development, are fertile and are able to nurse their young (data not shown).

Immunohistochemical staining for β-gal was performed on sections derived from entire inguinal mammary glands isolated from female $En\text{-}1^{Lki/+}$ transgenic mice, where *lacZ*-expressing cells produce a red-brownish reaction product. Importantly, no *lacZ* expression was detected in mammary glands derived from wildtype age-matched pubertal $En\text{-}1^{+/+}$ females (Fig. 2.2c).

At 28 d, where high levels of *En-1* expression were observed by western analysis, positive staining for β-gal was most obvious in putative myoepithelial and/or periductal fibroblast cells lining the extending ducts within mammary glands derived from $En\text{-}1^{Lki/+}$ females (see black arrows in Fig. 2.2d-f). *LacZ* expression was also detected in the cells presumably surrounding the tip of the highly proliferative TEB structures (see red

arrows in Fig. 2.2e,f) and possibly in nearby solitary stromal cells (see * in Fig. 2.2e). *En-1* expression appears to be restricted to the outermost layer of the TEB, where cap cells reside. At this early pubertal stage, where the growing epithelial tree is approaching the central lymph node, ß-gal expression was never identified in the lymph node or in luminal epithelial cells (Fig. 2.2d-f and data not shown). Of interest, ß-gal expression was no longer detectable in mammary tissue derived from 35-day old $En\text{-}1^{Lki/+}$ pubertal females (data not shown). In accordance with the temporal expression pattern presented in Fig. 2.1, *lacZ* expressing cells were never detected at timepoints earlier than day 15, later than day 35 or in subsequent developmental timepoints (data not shown), confirming the developmentally regulated expression of En-1 in the prepubertal and early pubertal mouse mammary gland.

_En-1_ deficient mammary glands exhibit impaired ductal growth and TEB formation

To overcome the neonatal lethality in *En-1* null mutants, the $En\text{-}1^{Lki}$ allele, which results in the complete disruption of En-1 function, was backcrossed into the C57Bl/6J background where rare $En\text{-}1^{hd/hd}$ null mutants have been reported to survive up to 3 weeks after birth [6]. The resulting $En\text{-}1^{Lki/+}$ heterozygous mice were indistinguishable from wildtype $En\text{-}1^{+/+}$ C57Bl/6J mice while $En\text{-}1^{Lki/Lki}$ null mutants were morphologically distinguishable at birth by their abnormal forelimb phenotype [7]. In addition, although $En\text{-}1^{Lki/Lki}$ null mutants were roughly the same size and weight as their siblings at birth, by 2-3 weeks of age they were noticeably smaller than their wildtype and heterozygous littermates (data not shown).

To determine the possible contribution of En-1 function to pubertal mammary gland development, entire inguinal mammary glands were surgically removed from rare surviving $En\text{-}1^{Lki/Lki}$ null mutant females, wholemounted and compared to mammary glands from wildtype or $En\text{-}1^{Lki/+}$ age-matched female littermates. Newborn and 21-day old $En\text{-}1^{Lki/Lki}$ female mammary glands resemble $En\text{-}1^{Lki/+}$ and $En\text{-}1^{+/+}$ female littermates (data not shown). At 28 days of age, TEBs are readily visible and the growing mammary

tree is approaching the lymph node in the early pubertal wildtype gland while no TEBs are visible and the gland does not appear to have grown as far into the underlying fat pad in the 28-day old *En-1*$^{Lki/Lki}$ null mutant littermate mammary gland (Fig. 2.3a, b). The TEBs are even more prominent and the ducts have now reached the central lymph node in a 32-day old *En-1*$^{Lki/+}$ mammary gland while the mammary glands from two *En-1* null mutant littermates are still devoid of TEBs, the ducts have yet to even approach the lymph node and the ductal tree still resembles a prepubertal epithelial tree (Fig. 2.3c-e).

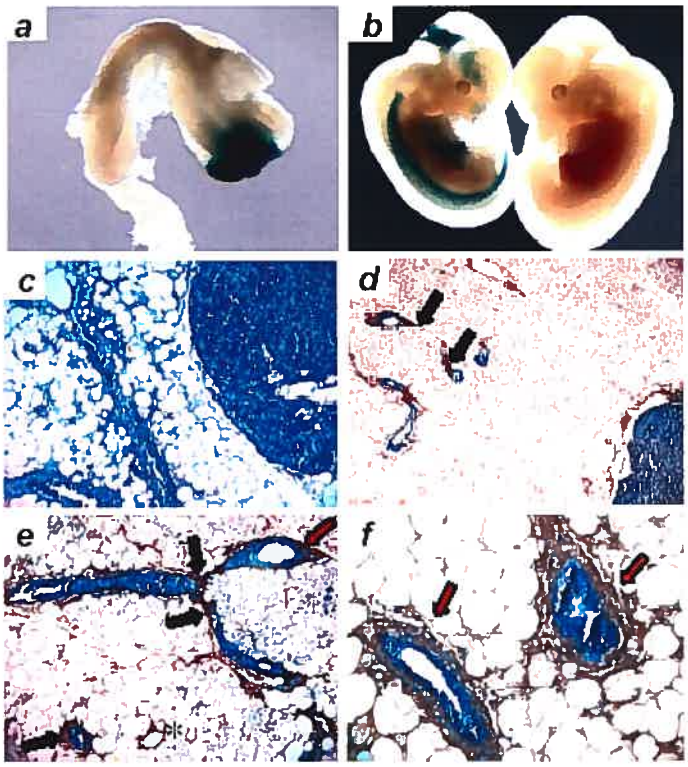

Fig. 2.2 *En-1* is expressed in duct-lining cells of the mouse mammary gland.
(a-b) E 8.5 and E12.5 dpc *En-1*$^{Lki/+}$ embryos stained *in toto* with X-Gal. Strong *lacZ*-derived activity is detected in the midbrain-hindbrain junction region (E8.5) and in the mid-hindbrain region, posterior hindbrain, spinal cord and somite-derived tissue (E12.5 embryo). (c-f) Immunohistochemical staining for *β*-gal expression on paraffin sections where "positive" cells are identified by the presence of a brownish reaction product in the presence of NovaRED substrate and sections are counterstained with hematoxylin

which stains nuclei blue. (c) Mammary glands from a representative 28-day old wildtype female shows absence of β-gal expression. (c-f) Mammary glands from 28-day old *En-1*$^{Lki/+}$ pubertal females show positive staining in epithelial cells surrounding the growing ducts (black arrows), the outer layer of TEBs (red arrows) and possibly in isolated stromal cells (see $*$ in e). Images were captured at either 10x (d), 20x (c,e) or 40x (f).

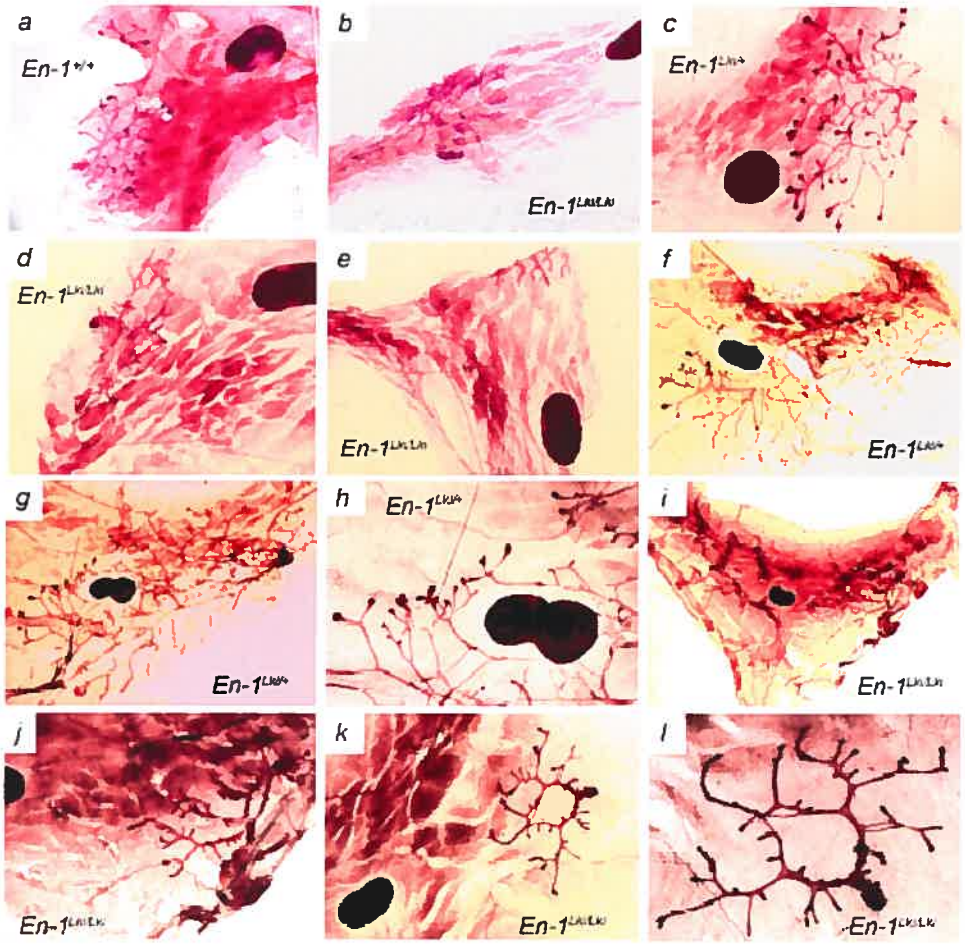

Fig. 2.3 Impaired ductal development in *En-1* null mutant mammary glands.
(a) Wildtype *En-1*$^{+/+}$ 28-day old mammary gland; mouse weight: 13.3 g. TEBs are readily visible and the growing mammary tree is approaching the lymph node at this stage of early puberty (1.6x). (b) Mutant *En-1*$^{Lki/Lki}$ 28-day old littermate mammary gland; mouse weight: 5.2 g. TEBs are not visible and the gland resembles a prepubertal gland (2x). (c) Control *En-1*$^{Lki/+}$ 32-day old mammary gland; mouse weight: 15.3 g. The

TEBs are visible and the ducts have reached the lymph node (1.6x). (d-e) Two mutant *En-1*$^{Lki/Lki}$ 32-day old littermate mammary glands; mouse weight: 6.1 and 5.3 g, respectively. TEBs are not visible (2x). (f-g) Control *En-1*$^{Lki/+}$ 42-day old mammary gland (left and right gland); mouse weight: 18.8 g. Some TEBs are still visible and the mammary tree has grown past the lymph node (0.8x). (h) Magnification of (g) showing the TEBs at the periphery of the growing ductal tree (2x). (i) Mutant *En-1*$^{Lki/Lki}$ 42-day old littermate mammary gland; mouse weight: 12.3 g. The mammary tree resembles a prepubertal tree and the TEBs are at best atrophic (0.8x). (j) Magnification of (i) (2x). (k) Contralateral mammary gland from the same *En-1*$^{Lki/Lki}$ mutant described in (j) (1.6x). (l) Magnification of (k) (3.2x).

Of all the mice analyzed in this study, one female *En-1*$^{Lki/Lki}$ homozygous mutant survived until 42 days of age and exhibited the most striking difference in mammary gland development due to *En-1* disruption. While the extensive ductal tree has grown past the lymph node and will soon completely fill the fat pad in the heterozygous *En-1*$^{Lki/+}$ female sibling (Fig. 2.3f-h), the *En-1*$^{Lki/Lki}$ homozygous null female revealed a primitive ductal system devoid of TEBs (Fig. 2.3i-l), reminiscent of the rudimentary mammary tree that is present upon birth. Consistent with the temporal expression pattern of *En-1*, where *En-1* is not expressed until day 15, these results suggest that while *En-1* is dispensable for the specification of the rudimentary tree present at birth, it may be required for the formation of TEB structures and the extension of the epithelial ductal system in preparation for sexual maturity. In addition, non-cell autonomous effects and/or a possible delay in sexual maturity, may contribute to, or exacerbate, this phenotype.

Reduced ductal morphogenesis in *En-1* null mammary epithelial transplants

To gain insight into whether the observed mammary gland phenotype in the *En-1* null mice was due to a cell-autonomous function of En-1 in mammary epithelium, control *En-1*$^{+/+}$, *En-1*$^{Lki/+}$ and mutant *En-1*$^{Lki/Lki}$ mammary epithelium were transplanted into wildtype syngeneic C57Bl/6J mammary fat pads following the removal of endogenous epithelium. In these studies, each 3 week old female recipient received an *En-1*$^{Lki/Lki}$

mammary tissue transplant in one inguinal fat pad and a control *En-1*$^{Lki/+}$ or *En-1*$^{+/+}$ transplant in the contralateral inguinal fat pad. Wholemount analysis of *En-1*$^{+/+}$ and *En-1*$^{Lki/+}$ mammary epithelial transplants harvested at 8 weeks (Fig 2.4 a,c,d) and 12 weeks (Fig. 2.4b) post transplantation show that the ducts were able to repopulate the host fat pad to full capacity and any remaining TEBs are located at the periphery of the fat pad before they regress, reminiscent of the epithelial tree seen in a normal postpubertal virgin mammary gland. In accordance with the fact that *En-1*$^{Lki/+}$ heterozygous mice have no mammary phenotype, *En-1*$^{Lki/+}$ mammary epithelial transplants undergo comparable ductal growth and morphogenesis as compared to the wildtype epithelium transplant (Fig. 2.4c-d). However, ductal growth and morphogenesis is severely affected in the *En-1* null mammary epithelial transplants as compared to wildtype or *En-1*$^{Lki/+}$ littermate transplants (Fig. 2.4e-j).

In contrast to the *En-1*$^{Lki/Lki}$ mutant mammary glands that resemble the primitive mammary tree present at birth, some of the *En-1* deficient mammary epithelium transplants, from 3 different homozygous *En-1*$^{Lki/Lki}$ mice, were able to grow and displayed impaired ductal outgrowth and branching morphogenesis to varying degrees when implanted into a wildtype fat pad microenvironment, from severe (Fig. 2.4e-g) to moderate (Fig. 2.4h-j). Two of the resulting *En-1*$^{Lki/Lki}$ transplants recapitulate the growth quiescent primitive rudimentary system seen in the *En-1* null mammary glands (Fig. 2.4e, g). Although one *En-1* null transplant appears to now be capable of forming TEB structures, the proliferative capacity of these structures is impaired since they fail to even penetrate the host fat pad beyond that of the mammary tree seen in a normal early prepubertal gland (Fig. 2.4f). Interestingly, some of the *En-1* deficient mammary epithelium transplants were able to expand into much of the fat pad but they show a reduction in branching and no visible TEBs when compared to wildtype or *En-1*$^{Lki/+}$ heterozygous reconstituted glands (e.g., compare Fig. 2.4i to d). In contrast to the well developed ductal network in the control *En-1*$^{+/+}$ and *En-1*$^{Lki/+}$ transplants, the ductal branching that emanated from the *En-1* null transplants did not extend to the periphery of the fat pad and failed to establish the same level of occupancy of the fat pad.

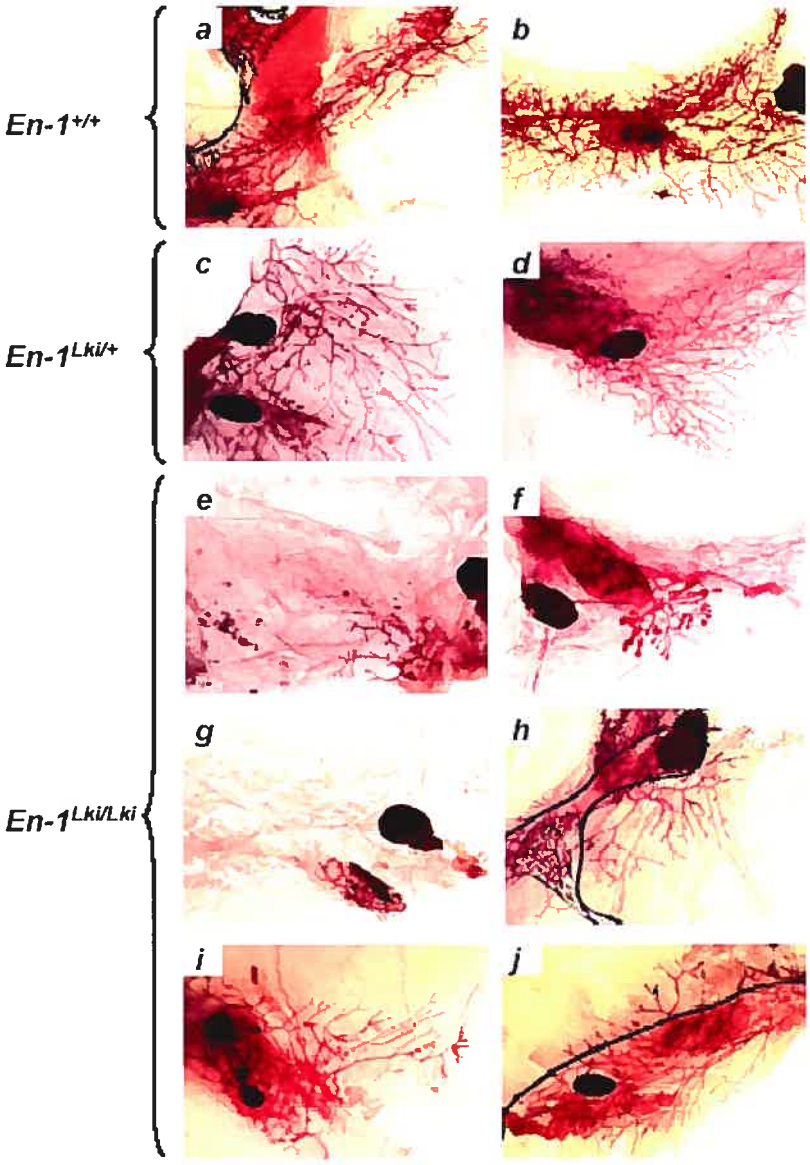

Fig. 2.4 Abnormal ductal development of transplanted *En-1* null mammary epithelium. (a,c-j) Wholemount analysis of mammary epithelial transplants of the indicated genotypes harvested at 8 or 12 (b) weeks post transplantation.

Discussion

In contrast to the paralog *En-2*, which is not expressed at any developmental timepoint

in the normal mammary gland, this study demonstrates the unique spatio-temporal pattern of *En-1* expression during mammary gland development in the young female mouse. *En-1* expression is highly regulated and is limited to the prepubertal gland and during the onset of puberty. Immunohistochemical analysis of $En\text{-}1^{Lki/+}$ mammary tissue suggests that *En-1* is expressed in epithelial cells lining the expanding ductal system and in the proliferative cap cells of TEBs. Based upon wholemount analysis, mammary gland development was equivalent at birth in $En\text{-}1^{Lki/Lki}$, $En\text{-}1^{Lki/+}$ and wildtype littermates, suggesting that prenatal mammary development is normal in *En-1* homozygous null mice. Rare mutant $En\text{-}1^{Lki/Lki}$ females that survived to different pubertal stages revealed a primitive ductal rudiment with a dramatic reduction in TEBs while a fully developed ductal tree was seen in the wildtype siblings. Interestingly, a significant proportion of the *En-1* null mammary epithelium transplants were able to expand into the wildtype fatty stroma but displayed limited ductal growth and reduced side branching to different degrees in the normal microenvironment and hormonal milieu of the wildtype host fat pad.

The phenotype observed in the *En-1* homozygous null female gland is not entirely reflective of a mammary epithelial autonomous defect since it was not consistently recapitulated by transplantation into wildtype hosts. At least within the scope of this limited analysis, the transplantation experiments show that the *En-1* homozygous null mammary epithelium undergoes ductal extension, albeit somewhat reduced, if given the correct developmental cues from the wildtype stromal microenvironment. Given that the rare *En-1* homozygous null mice that survive fail to thrive, it is thus conceivable that the hormonal milieu of these mice is far from optimal. Moreover, and not tested here, it is possible that En-1 may respond to hormonal changes that occur at the onset of puberty.

The second possible non cell-autonomous defect in these mice could result from a defective microenvironment. Indeed, we detected rare β-gal positive non-epithelial cells in the vicinity of the expanding pubertal mammary epithelial tree (see * in Fig. 2.2e). However, western blot analyses identified En-1 protein expression only in portions of mammary tissue that included mammary epithelial fragments, and not in the adjacent fat

pad, suggesting that the stromal compartment does not contribute to the observed phenotype. Reciprocal experiments (i.e., transplanting wildtype cells into cleared fat pads from En-1 mutant mice) could help clarify this possibility but, due to viability issues, their feasibility is compromised in homozygous *En-1* null mice. Thus, the significance of this observation (i.e., possible expression of *En-1* in rare non-epithelial cells) remains unclear and may be best appreciated using experimental strategies that will complement those exploited in our study. For example, conducting recombinant epithelium/stroma capsular kidney cultures as described by Wiesen et al. (1999) may be very informative [8].

Although there is an obvious non-cell autonomous contribution to the observed mammary gland phenotype, it is also likely that the consistent reduction in ductal density in the subgroup of glands reconstituted with cells lacking a functional *En-1* gene also involves cell-autonomous effects. In agreement with the predominant expression of En-1 in cells lining the epithelial ducts, the reduction in ductal tree density observed in the reconstituted glands more likely reflects the possibility that *En-1* expression in the mammary epithelium is required for epithelial-mesenchymal interactions that condition the stroma such that it allows or stimulates the mammary epithelium to grow. The limited number of biological samples (n=3) precluded us from performing a meaningful quantitative morphometric analysis of total mammary epithelial structures to stromal area as performed by Wiesen et al. (1999) [8]. Further studies, where the possible variable of an overall delay in development is removed, would be most valuable. Based on the present study alone, it remains difficult to quantify the contribution of cell-autonomous versus non-cell autonomous effects.

It is possible that the modifying genes in the C57Bl/6J strain that temper the normally lethal brain phenotype also temper the capacity of mammary epithelium to grow upon transplantation. This might explain why C57Bl/6J *En-1* null mammary epithelium is able to form TEBs and grow when transplanted into a C57Bl/6J recipient, a less severe mammary gland phenotype than what is observed in the *En-1* null mammary gland. It is also possible that compensatory ectopic expression of *En-2* in $En\text{-}1^{Lki/Lki}$ C57Bl/6J mice

may partially rescue the mammary gland phenotype in these mutant mice, analogous to the finding that *En-2* expression is likely responsible for the less severe cerebellar brain phenotype when the *En-1 Lki* knock-in gene is introduced into the C57Bl/6J strain [6]. Although we have determined that *En-2* expression is never detected in mammary tissue from both CD-1 and C57Bl/6J strains (Fig. 2.1 and data not shown), *En-2* expression was never tested in the actual rare $En\text{-}1^{Lki/Lki}$ mutants that arose.

Given that the loss of En-1 function in the mouse mammary gland leads to reduced growth in the otherwise rapidly proliferating pubertal ductal system, it is interesting to note that ectopic expression of *EN2* in human breast tissue is associated with breast cancer [3]. Since the consequence of the loss of En-1 function in the mammary gland is consistent with a role for *En-1* in regulating proliferation, combined with the fact that ectopic expression of *En-1* in mammary epithelial cells promotes proliferation, it is tempting to speculate it plays a similar role in normal mammary tissue and that the aberrant expression of *EN2* in mammary tissue leads to the inappropriate activation of En-1 targets that promote cell proliferation [3]. It will therefore be important to identify if there are common key target genes that are both regulated by En-1 in the normal expansion of the mammary gland and activated by ectopic *EN2* expression in human breast cancer.

Materials and Methods

Transgenic mice

The generation of $En\text{-}1^{Lki}$ (*Lki, lacZ* knock-in) transgenic mice, where the bacterial gene *lacZ* which codes for *ß*-galactosidase (*ß*-gal), was inserted into the *En-1* locus, has been reported previously [5]. The insertion of the bacterial *lacZ* coding sequences into the first exon of *En-1* brought the resulting expression of *ß*-galactosidase *(ß*-gal*)* under the regulation of the *En-1* promoter and endogenous regulatory elements. In $En\text{-}1^{Lki/+}$ mice heterozygous for the targeted allele, the profile of *ß*-gal activity during embryogenesis was shown to accurately recapitulate that of the corresponding *En-1* transcripts, as determined by double-labeling using X-gal to detect *lacZ* expression and an antibody to

detect En-1 protein and thus serves as a faithful reporter of *En-1* expression [9]. The first 111 amino acids of *En-1* are deleted by the above insertion of *lacZ*, and the resulting *En-1 Lki* allele results in the complete disruption of En-1 function as the homozygous *En-1*$^{Lki/Lki}$ mutants are morphologically identical to null *En-1*$^{hd/hd}$ mutants (*hd*, deletion of the homeodomain, also resulting in the complete disruption of En-1 function) and En-1 protein expression is not detected with anti-En antibodies [9]. The CD1 *En-1*$^{Lki/+}$ mice were backcrossed into and maintained on a C57Bl/6J background where rare *En-1*$^{Lki/Lki}$ mutants survive 3-4 weeks after birth [6]. *En-1*$^{Lki/+}$ heterozygous mice were indistinguishable from wildtype *En-1*$^{+/+}$ mice while *En-1*$^{Lki/Lki}$ null mutants were morphologically distinguishable by their forelimb phenotype [7]. *En-1*$^{Lki/+}$ and *En-1*$^{Lki/Lki}$ mice were genotyped by Southern blot analysis of EcoRI-digested tail DNA using a 2400-bp Xho l-Sst l fragment from the bacterial *lacZ* cDNA and/or PCR analysis. Primers used for genotyping the *En-1* wildtype allele were 5'-GAA AAA AGA AAG GCG AGC GTC-3' and 5'-AGC CTA AAA GTC AGC GCG AC-3', which produced a 209 bp product. Primers used for genotyping the *En-1*^{Lki} allele were 5'-AAT CCC GAA TCT CTA TCG TGC-3' and 5'-CAC TCG GGT GAT TAC GAT CG-3', which produced a 240 bp fragment. All animals were maintained and bred in ventilated microisolator cages, provided with sterilized food and acidified water in the specific pathogen-free (SPF) animal facilities of the Clinical Research Institute of Montreal (IRCM) and of l'IRIC.

Cell lines

The HC11 mammary epithelial cell line is a clonal derivative of the COMMA-1D cell line, derived from mammary tissue of a mid-pregnant BALB/c female [10]. The C57MG cell line was derived from the glands of a 23 week old retired C57BL/6 breeder [11]. The MCF7 mammary epithelial cell line was derived from a human breast adenocarcinoma. C57MG, HC11 and MCF7 lines were grown in RPMI 1640 medium supplemented with 10% fetal calf serum, 10 ng/ml of epidermal growth factor and 10 µg/ml of insulin. The generation of HC11 and C57MG mammary epithelial cells engineered to express *En-1* have been described previously [3].

Western Blot Analysis

Total protein was extracted from frozen #4 (inguinal) mouse mammary glands, adult mouse cerebellum and cell lines and 40 µg of each resulting sample was separated by SDS-PAGE as described previously [12]. En proteins were detected with *αEnhb-1* antisera (which detects both 41 kDa mouse En-1 and human EN1, and both 55 kDa mouse En-2 and human EN2 proteins) as described [13]. As a control for loading, membranes were stripped and hybridized with either *αPTP1D* (protein-tyrosine phosphatase 1D; P54420; BD PharMingen, Mississauga, Canada) or *αTubulin* (mouse monoclonal to beta Tubulin; ab7287; Abcam, Cambridge, MA).

ß-galactosidase staining of embryos and immunohistochemical detection of *lacZ* expression

Embryos were removed from *En-1*$^{Lki/+}$ females, fixed in 4% paraformaldehyde (PFA) overnight on a rotisserie at 4°C, and stained *in toto* for ß-galactosidase (ß-gal) activity as described [14]. For ß-gal detection within the mouse mammary gland, entire inguinal mammary glands were removed from *En-1*$^{Lki/+}$ females at the indicated ages and fixed in chilled 2% PFA, 0.2% glutaraldehyde, 0.02% Nonidet P-40 in PBS for 2 hours and subsequently thoroughly rinsed in 4 changes of chilled PBS before being dehydrated, paraffin-embedded and sectioned at 5 µm. Immunohistochemical detection of *lacZ* expression was performed on the resulting dewaxed and rehydrated sections using the Vectastain Elite ABC Kit (PK-6101; Vector Laboratories, Burlingame, CA). Tissue sections were incubated with a rabbit anti-E.coli-ß-galactosidase polyclonal antibody for 1 hour in a moisture chamber at RT (1:500 dilution; ab616; Abcam Inc., Cambridge, MA). After PBS washes, sections were incubated with biotinylated goat anti-rabbit IgG before incubation with ABC (avidin-biotinylated-enzyme-complex) conjugated horse-radish peroxidase (PK-6101; Vector Laboratories, Burlingame, CA). Peroxidase was visualized by incubating sections with NovaRED substrate (SK-4800; Vector Laboratories, Burlingame, CA) and slides were counterstained with hematoxylin.

Wholemounts and mammary gland transplantation

For wholemount analysis, entire inguinal mammary glands were surgically removed at the indicated timepoints and spread flat on glass slides. After fixation in Carnoy's fixative for 2 hours, the glands were defatted in ethanol and acetone washes and then rehydrated and stained in Carmine Alum Stain overnight. The glands were then dehydrated, cleared with xylene and mounted with Permount. For *En-1* null mammary gland transplantation studies, rare surviving *En-1*$^{Lki/Lki}$ mutant female C57Bl/6J mice were identified shortly after birth by their limb phenotype and were left to thrive with their female *En-1*$^{+/+}$ and/or *En-1*$^{Lki/+}$ littermates until 3 weeks of age. At 3 weeks of age, the entire prepubertal mammary epithelial tree within the inguinal mammary gland (the portion of the mammary gland comprising the main duct, rudimentary ductal system and the surrounding fatty stroma from the nipple until the middle of the lymph node) was removed from *En-1*$^{Lki/Lki}$ mutant females, *En-1*$^{+/+}$ and/or *En-1*$^{Lki/+}$ female littermates and wildtype 3 week old syngeneic *En-1*$^{+/+}$ C57Bl/6J female mice. Using scissors, a small incision was made in the remaining epithelial-free fat pad portion of the age-matched *En-1*$^{+/+}$ C57Bl/6J recipient gland. Using dissection tweezers, the excised fat pad portion from the *En-1*$^{Lki/Lki}$ female was placed into this incision within the cleared fat pad of the syngeneic host, just above the lymph node. The surgical procedures for clearing the endogenous mammary epithelium from the inguinal fat pads of 3 week-old female mice and implanting tissue fragments into cleared mammary fat pads have been described [2]. Each recipient female carried an *En-1*$^{Lki/Lki}$ mutant mammary tissue transplant in one fat pad and a control *En-1*$^{Lki/+}$ or *En-1*$^{+/+}$ transplant in the contralateral fat pad. Wholemount preparations of the excised portion of the host mammary gland were generated and examined to ensure that only those *En-1*$^{+/+}$ transplant recipients where the entire mammary epithelial tree was removed were included in the study. Tissue was reserved from each sacrificed animal to confirm the genotype of mutant *En-1*$^{Lki/Lki}$ transplants, female *En-1*$^{+/+}$ or *En-1*$^{Lki/+}$ littermate transplants and recipient *En-1*$^{+/+}$ C57Bl/6J females. The female hosts were sacrificed 8-12 weeks post-transplantation and the chimeric mammary gland transplants were surgically removed and examined by wholemount analysis.

Acknowledgements

We thank Dr. Alexandra Joyner (NYU, NYC) for *En-1* and *En-2* cDNAs and α*Enhb-1* antisera; Dr. Shiu (University of Manitoba, Winnipeg) for the MCF7 cell line; Dr. Paul Jolicoeur (IRCM) for the HC11 cell line; Dr. Anthony Brown (Cornell University, NYC) for the C57MG cell line; Melanie Frechette for assistance in the animal facility, Christian Charbonneau for assistance in photography, and Denis Rodrigue for histological sectioning services. This work was supported by funds from the NCIC (15064) to G.S. N.M. was a recipient of a CIHR research award. G.S. is a scholar of the leukemia and lymphoma society of America and a recipient of a Canada Research Chair in Molecular Genetics of Stem Cells.

References

1. Hennighausen,L. & Robinson,G.W. Think globally, act locally: the making of a mouse mammary gland. *Genes Dev.* **12**, 449-455 (1998).

2. DeOme,K.B., Faulkin,L.J., Bern,H.A. & Blair,P.B. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *J.Natl.Cancer Inst.* **78**, 515-525 (1959).

3. Martin,N.L. et al. *EN2* is a candidate oncogene in human breast cancer. *Oncogene.* (2005).

4. Wurst,W., Auerbach,A.B. & Joyner,A.L. Multiple developmental defects in Engrailed-1 mutant mice: an early mid-hindbrain deletion and patterning defects in forelimbs and sternum. *Development.* **120**, 2065-2075 (1994).

5. Hanks,M. et al. Rescue of the En-1 mutant phenotype by replacement of En-1 with En-2. *Science.* **269**, 679-682 (1995).

6. Bilovocky,N.A. et al. Factors in the genetic background suppress the engrailed-1 cerebellar phenotype. *J.Neurosci.* **23**, 5105-5112 (2003).

7. Loomis,C.A. et al. The mouse Engrailed-1 gene and ventral limb patterning. *Nature.* **382**, 360-363 (1996).

8. Wiesen,J.F., Young,P., Werb,Z. & Cunha,G.R. Signaling through the stromal epidermal growth factor receptor is necessary for mammary ductal development. *Development.* **126**, 335-344 (1999).

9. Matise,M.P. & Joyner,A.L. Expression patterns of developmental control genes in normal and Engrailed-1 mutant mouse spinal cord reveal early diversity in developing interneurons. *J.Neurosci.* **17**, 7805-7816 (1997).

10. Ball,R.K. et al. Prolactin regulation of beta-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. *EMBO J.* **7**, 2089-2095 (1988).

11. Vaidya,A.B., Lasfargues,E.Y., Sheffield,J.B. & Coutinho,W.G. Murine mammary tumor virus (MuMTV) infection of an epithelial cell line established from C57BL/6 mouse mammary glands. *Virology.* **90**, 12-22 (1978).

12. Thorsteinsdottir,U. et al. The oncoprotein E2A-Pbx1a collaborates with Hoxa9 to acutely transform primary bone marrow cells. *Mol.Cell Biol.* **19**, 6355-6366 (1999).

13. Davis,C.A., Holmyard,D.P., Millen,K.J. & Joyner,A.L. Examining pattern formation in mouse, chicken and frog embryos with an En-specific antiserum. *Development*. **111**, 287-298 (1991).

14. Logan,C., Khoo,W.K., Cado,D. & Joyner,A.L. Two enhancer regions in the mouse En-2 locus direct expression to the mid/hindbrain region and mandibular myoblasts. *Development*. **117**, 905-916 (1993).

CHAPITRE 3

ARTICLE

EN2 is a Candidate Oncogene in Human Breast Cancer

Nicole L. Martin[1], Marc K. Saba-El-Leil[4], Svetlana Sadekova[5], Sylvain Meloche[4] and Guy Sauvageau[1,2,3 *]

Published online June 20, 2005

Oncogene, (2005) 24(46):6890-901.

[1]Laboratory of Molecular Genetics of Hematopoietic Stem Cells, Institut de Recherches en Immunovirologie et Cancérologie, C.P. 6128, succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada; [2]Department of Medicine and [3]Division of Hematology, Maisonneuve-Rosemont Hospital, Montréal, Québec, Canada, [4]Laboratory of Signaling and Cell Growth, Institut de Recherches en Immunovirologie et Cancérologie, C.P. 6128, succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada [5]Breast Cancer Functional Genomics Group, McGill University, 687 Pine Ave West, Montréal, Québec H3A 1A1.

Chapter 3 is a co-authored article which documents the ectopic expression of *EN2* in human breast-tumor derived cell lines and primary human breast cancer samples. Most importantly, this work was the first to investigate a possible role of *EN2* in mammary tumorigenesis and describes *in vitro* and *in vivo* studies which suggest that *EN2* indeed behaves as an oncogene in mammary tissue.

Dr. Marc Saba-El-Leil, a Research Associate in the laboratory of Dr. Sylvain Meloche at the time, contributed to this work by performing the transfections of synthetic siRNA using a transfection procedure developned in the laboratory of Dr. Meloche. Dr. Svetlana Sadekova, who heads the Breast Cancer Functional Genomics Group within the Oncology Group at McGill University, contributed to this work by provding the 11 frozen primary normal human reduction mammoplasty breast tissue samples presented in Figure 3.2, the frozen primary human infiltrating ductal carcinoma, inflammatory type sample represented by T4 in Figure 3.3 and the additional 59 primary human breast tumor samples tested (data not shown). The paper was written by Nicole Martin under the supervision of Dr. Guy Sauvageau.

Abstract

Only a few critical oncogenes have been identified in the more commonly occurring cases of sporadic breast cancer. We provide evidence that *EN2* is ectopically expressed in a subset of human breast cancer and may have a causal role in mammary tumorigenesis. Non-tumorigenic mammary cell lines engineered to ectopically express *En-2* have a marked reduction in their cycling time, lose cell contact inhibition, become sensitive to 17-AAG treatment, fail to differentiate when exposed to lactogenic hormones and induce mammary tumors when transplanted into cleared mammary glands of syngeneic hosts. RNA interference studies suggest that *EN2* expression is required for the maintenance of the transformed phenotype of a human breast tumor cell line.

Introduction

Breast cancer is one of the most frequent human malignancies in the Western world. The pathogenesis of this disease is thought to involve multiple genetic and epigenetic events. In spite of recent advances in the assessment of breast cancer risk, through the identification of crucial susceptibility genes (*BRCA1/2, PTEN*, *P53*), these account for less than 5% of all breast cancer cases and may not be associated with the more commonly occurring sporadic breast cancers [1]. The discovery of bona fide primary genetic lesions underlying sporadic breast cancer development remains a major challenge. This is due, at least in part, to the marked cytogenetic complexity seen in advanced breast cancer, precluding investigators from readily identifying primary causative genetic events in breast cell transformation [2].

A number of oncogenes and tumor suppressors have been associated with breast cancer. The *c-MYC* gene is amplified and/or overexpressed in a high proportion of human breast cancer, although the frequency of these alterations varies greatly [3]. *ErbB2* is also amplified and subsequently overexpressed in 20-30% of human breast cancers, and overexpression of *ErbB2* is correlated with a poor clinical prognosis of both node-positive and node-negative tumors [4]. The *Cyclin D1* gene is amplified in 15-20% of human breast cancers [5]. As the oncogenes located at amplified chromosomal regions are rarely amplified in benign breast disease [6], they may represent later events in the multistep progression associated with the development of breast cancer.

In rodents, it has been possible to identify several breast cancer-initiating oncogenes by the characterization of proviral integration sites of the mouse mammary tumor virus (MMTV). One such locus is *Wnt-1* which, when overexpressed, leads to mammary hyperplasia and subsequent generation of adenocarcinomas [7]. Although expression of *WNT-1* itself has not been reported in normal or neoplastic human breast tissue, other *WNT* genes have been detected in subsets of human breast cancers [8]. *Wnt-1* encodes a secreted growth factor that initiates a signaling cascade which results in transcriptional activation mediated by β-catenin/Tcf complexes [9]. β-catenin/Tcf-mediated transcription has also been implicated in human cancer, with some targets relevant to breast

carcinogenesis identified such as c-*MYC* and *cyclin D1* [10,11]. The observation that *Wnt-1* is a mouse mammary oncogene and that the downstream mediator β-catenin is often stabilized in certain human malignancies fuels the ongoing search for additional targets of this pathway, such as the *Engrailed* genes, that might also be implicated in breast cancer.

The mouse *En-1* and *En-2* genes encode homeobox-containing transcription factors that are the murine homologs of the *Drosophila* segment polarity gene *engrailed. En-1* is first expressed in the presumptive mid/hindbrain around 8.0 dpc and continues to be expressed, together with *En-2*, in overlapping patterns during midbrain development [12]. *En-2* expression is restricted to the central nervous system and branchiolar arches during embryogenesis and *En-2* null mutants are viable but harbor reductions in cerebellar size [13,14].

En and other homeobox-containing genes clearly occupy a prominent position in the developmental regulatory hierarchy, yet they have received little attention with respect to mammary gland organogenesis and cancer. In the present study we show that *EN2* (but not *EN1*) is ectopically expressed in a subset of human breast cancer and in a large proportion of breast cancer cell lines and that its ectopic expression readily transforms mammary epithelial cells *in vitro* and promotes adenocarcinoma formation *in vivo*. We also provide evidence of its critical function in a breast cancer cell line.

Results

Engrailed genes are rarely activated in Wnt-1-induced mammary tumors

As *Engrailed-1* (*En-1*) and *Engrailed-2* (*En-2*) are functional targets of *Wnt-1* in mouse embryogenesis and several different groups of homeodomain proteins have been implicated as causative oncogenes in cancer [15], we initially sought to determine whether *En-1* or *En-2* were implicated in Wnt-1-induced mouse mammary hyperplasia observed in MMTV-Wnt-1 transgenics [16]. Neither *En-1* nor *En-2* were detected by western blot or RT-PCR analyses in hyperplastic mammary glands derived from nulliparous

hemizygous MMTV-Wnt-1 transgenic females (n=4; Fig. 3.1a, lane 2,3,5,7). In addition, only one of three tumors arising from the MMTV-Wnt-1 transgenics expressed *En-1* (not shown), while no expression of *En-2* was detected in these primary tumors (Fig. 3.1a). Thus, while *En-1* and *En-2* are downstream and responsive to Wnt-1 signaling in embryogenesis, it seems unlikely that they contributed to tumor formation.

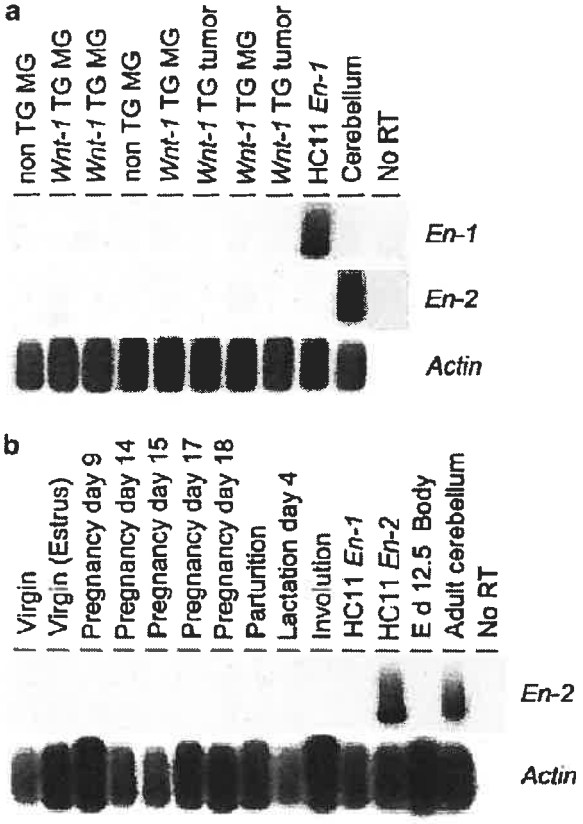

Fig. 3.1 Expression of *En-1* and *En-2* in the mouse mammary gland. (a) Semi-quantitative RT-PCR analysis of globally amplified cDNA derived from hyperplastic mammary glands (MG) and tumors of MMTV-Wnt-1 transgenic mice. (b) RT-PCR analysis was used to investigate *En-2* expression at several different developmental stages in the mouse mammary gland. HC11 mammary epithelial cells engineered to express *En-1* or *En-2,* E12.5 dpc embryonic bodies without heads (Ed12.5 Body), where

En-1 is exclusively expressed, and adult mouse cerebellum, where *En-2* is exclusively expressed, were included as specificity controls.

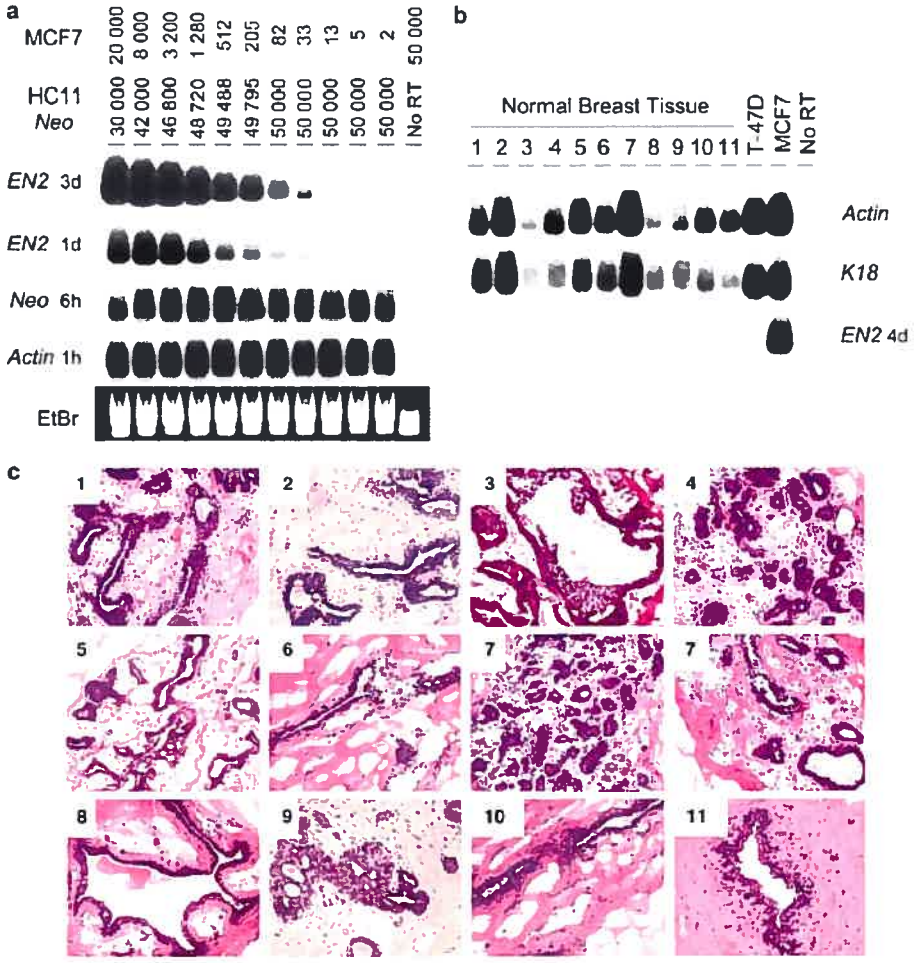

Fig. 3.2 *EN2* is not detectable in epithelial structures within normal human breast tissue. (a) *EN2* expression in total amplified cDNA obtained from consecutive 2.5 fold dilutions of MCF7 cells (which express *EN2*) with HC11 *Neo* cells (*EN2* negative). (b) RT-PCR analysis of 11 normal human breast tissue samples. *Keratin 18* was used to show the epithelial content in each sample. (c) H&E sections from the corresponding 11 samples in (b) were taken both above and below the section used to isolate RNA (20x).

Although *En-1* shows a specific temporal expression pattern in the normal adult mouse mammary gland (NM and GS, manuscript in preparation), *En-2* expression was not detectable in the normal mouse gland (Fig. 3.1b). With a detection sensitivity nearing 0.01% of the cells (Fig. 3.2a), *EN2* was also not detected in normal human mammary tissues derived from reduction mammoplasty (n=11; Fig. 3.2b and c). As detailed below however, the presence of EN2 protein in the human breast cancer cell line MCF7 (Fig. 3.3a) led us to investigate its expression in primary human breast tumors.

_EN2_ is ectopically expressed in human breast cancer samples

Using semi-quantitative RT-PCR analysis, we found that *EN2* expression was not limited to MCF7 cells since a large proportion (7/12 or 58%) of established breast carcinoma cell lines expressed this gene (Fig. 3.3b). Of the *EN2*-positive cell lines, four were derived from adenocarcinomas (MDA-MB-435S, BT-20, MDA-MB-436 and MCF7), one from ductal carcinoma (BT-474), and two were designated as fibrocystic breast tissue (MCF 10A and MCF-12A). The latter two lines, although originally derived from hyperplastic breast tissue and not carcinomas, have evolved from the initial immortalized normal cell line as they now form colonies in soft agar, a characteristic frequently associated with transformed cells (data not shown). Western blot analyses confirmed the presence of EN2 and the absence of EN1 in all of these cell lines (data not shown).

To determine whether *EN2* was also expressed in primary human breast tumors, semi-quantitative RT-PCR analysis was performed on RNA derived from frozen breast biopsies. Two of the 23 primary tumors initially analyzed expressed *EN2* at levels comparable to that observed in MCF7 cells (Fig. 3.3c). Importantly, the normal tissue adjacent to these two tumors did not express *EN2* (Normal 3 and 4; Fig. 3.3c, lane 1 and 16, respectively). These observations were confirmed by immunohistochemistry which revealed strong EN2 nuclear staining in the neoplastic epithelial cells within the tumors only and not in the normal adjacent tissue (Fig. 3.3d, compare T3 and T4 to N3 and N4, respectively). Histologically, these 2 tumors were defined as adenocarcinoma (T3) and

infiltrating ductal carcinoma, inflammatory type (T4). Our analysis of human breast tumors and normal adjacent samples was expanded to better define an estimated frequency of *EN2* expression in breast cancer. While *EN2* was not detected in any of the normal samples analyzed (n=19 total), 4/59 of the additional tumors expressed *EN2* (not shown). All four additional *EN2* positive tumors were infiltrating ductal carcinomas. Thus, *EN2* is ectopically expressed in a subset of human breast cancer representing approximately 7% in the population tested.

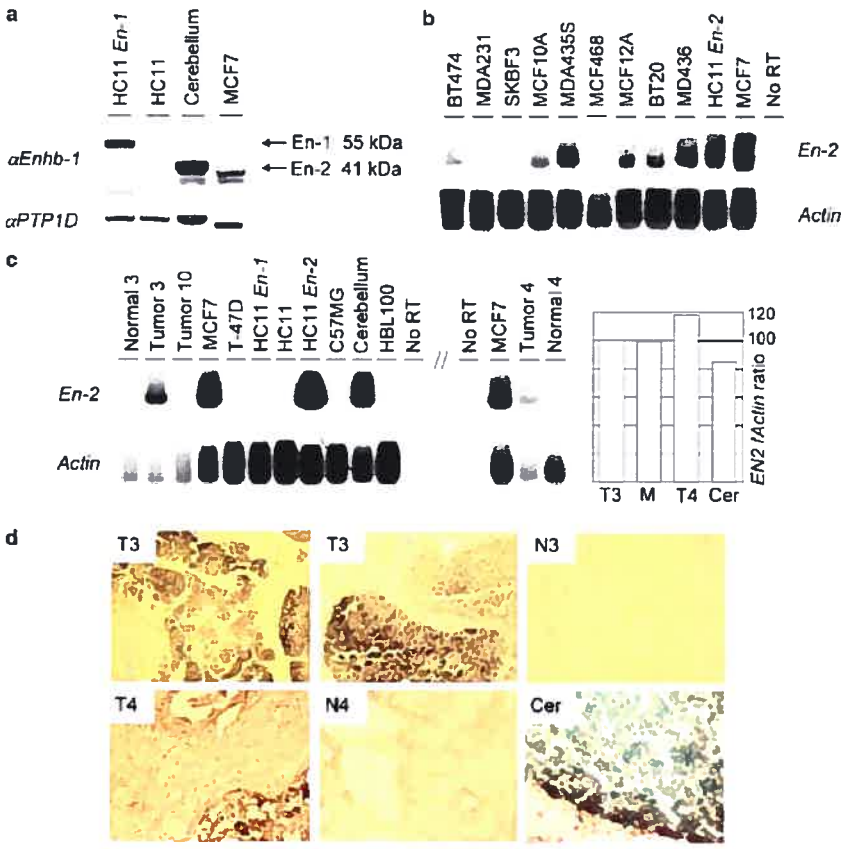

Fig. 3.3 *EN2* is ectopically expressed in human breast carcinomas. (a) Western blot analysis shows the presence of EN2 in the human breast cancer cell line MCF7. *αEnhb-1*; antiserum that recognizes EN1 and EN2, *αPTP1D*; protein-tyrosine phosphatase 1D antibody. (b) RT-PCR analysis of RNA isolated from human breast tumor-derived cell lines. (c) RT-PCR analysis of RNA derived from primary breast tumor biopsies shows

ectopic *EN2* expression in two tumors (Tumor 3 and 4), while *EN2* expression is not detected in the normal adjacent tissue to each tumor (Normal 3 and 4). Quantitative phosphoimager analysis of the *EN2 / Actin* ratio in Tumor 3 (T3), Tumor 4 (T4), MCF7 (M), and endogenous levels in the adult mouse cerebellum (Cer) in the figure are shown. (d) Immunohistochemistry using *αEnhb-1* antisera on sections derived from Tumor 3 (20x); in the normal adjacent tissue to Tumor 3 (20x); in the Tumor 4 section (40x); and in the normal adjacent tissue to Tumor 4 (40x); *En-2* expression in the adult mouse cerebellum has been documented and was used as a positive control (40x).

Ectopic expression of *En-2* readily transforms mammary epithelial cell lines

To investigate the possible oncogenic role of *En-2* in normal breast epithelial cells, we first engineered, by retroviral gene transfer, its ectopic expression in two immortalized, non-transformed, anchorage-dependent mammary epithelial cell lines, HC11 and C57MG. HC11 cells were chosen as they have retained both the ability to differentiate *in vitro* upon stimulation with lactogenic hormones and to generate epithelial outgrowths when transplanted back into the cleared (gland-free) fat pads of syngeneic hosts, as observed with primary mammary epithelial cells [17]. Additionally, both HC11 and C57MG cells have been shown to acquire anchorage independent growth when transduced with oncogenes involved in breast cancer (*c-erbB-2* in HC11:[18]; *Wnt-1* in C57MG:[19]). The resulting levels of *En-2* expression in HC11 selected polyclonal populations transduced with *En-2*-containing retrovirus were comparable to levels of *EN2* seen in MCF7 and endogenous levels seen in the cerebellum (Fig. 3.4).

HC11 and C57MG cells ectopically expressing *En-2* proliferated significantly faster when compared to parental untransduced cells or *Neo*-transduced cells (Fig. 3.5a and b). Furthermore, ectopic *En-2* expression conferred anchorage independent growth to both cell lines (Fig. 3.5c). *En-2*–dependent loss of cell contact inhibition was also observed in HC11 cells (Fig. 3.5d). All of the above effects produced by *En-2* expression, were reproduced, albeit to a lesser extent, with the paralogous gene *En-1* (Fig. 3.5).

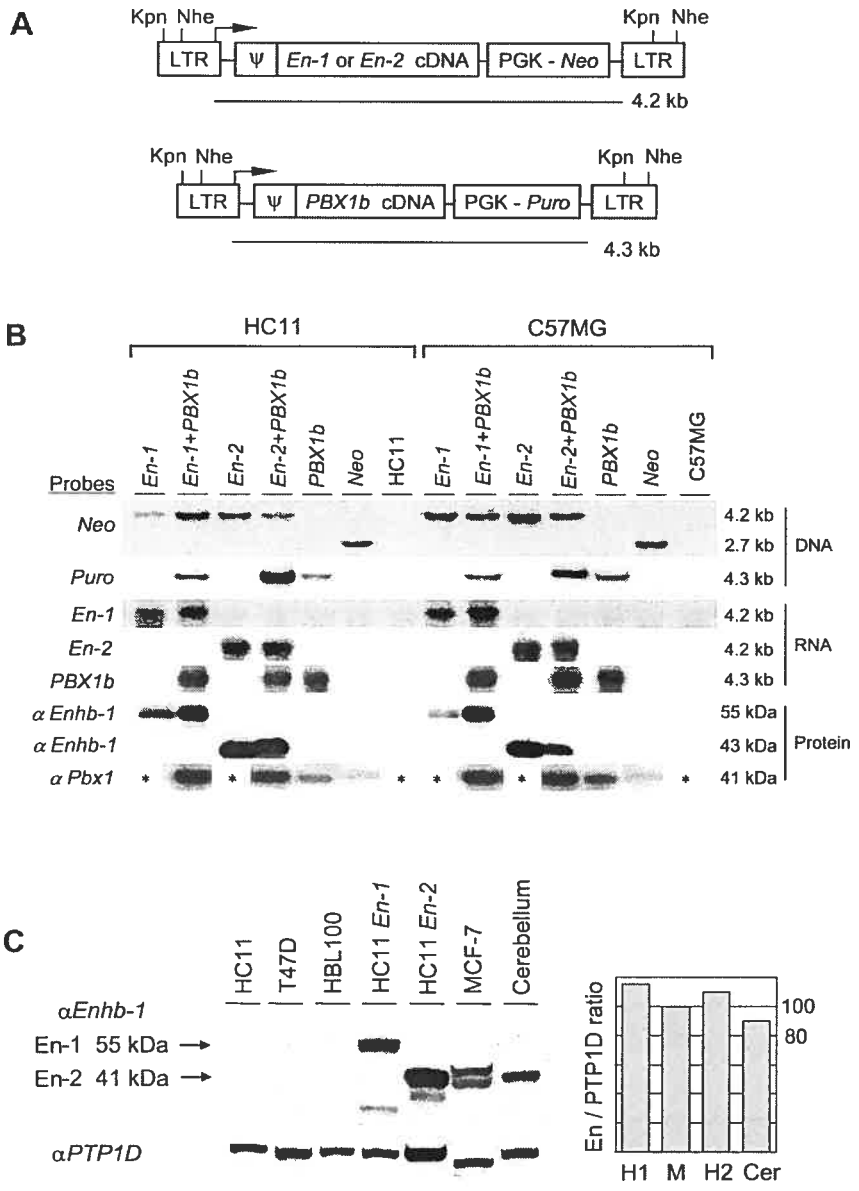

Fig. 3.4 HC11 and C57MG cells engineered to ectopically express *En-2* by retroviral gene transfer. (a) Schematic representation of the retroviral constructs carrying the full length cDNA for *En-1*, *En-2* or *PBX1b*. The expected sizes of the LTR driven viral transcripts are shown. (b) The integrity and expression of the *En-1, En-2, PBX1b* and *Neo*-containing proviruses within the selected polyclonal transduced populations were confirmed by Southern (DNA), northern (RNA) and western blot analysis (Protein). The genomic DNA was digested with either Kpn I or Nhe I to release the integrated *En-1/En-*

2 (4.2 kb), *PBX1b* (4.3 kb) and *Neo* (2.7 kb) viruses where *Neo*- and *Puro*-specific cDNA probes detect the *En-1/En-2* and *Neo*, and *PBX1b* proviruses, respectively. *En-1/En-2 and PBX1b* transcripts are detected in total RNA from the same cell populations. Autoradiographs were exposed 14 hr at -70°C. Western blot analysis using α*Enhb-1* antiserum confirms that both En-1 (55 kDa) and En-2 (41 kDa) are present at the protein level in cells transduced with either *En-1 or En-2*-containing retrovirus. (c) Western blot analysis and the corresponding En / PTP1D ratio of En-1 (H1) and En-2 (H2) protein levels achieved by retroviral gene transfer in comparison to that documented in MCF7 (M) and endogenous levels seen in the adult mouse cerebellum (Cer).

Similar to Hox proteins, Engrailed proteins can bind target DNA as a heterodimer with Pbx1b, another homeodomain-containing protein [20] [21]. Studies in our lab have shown that Hox-induced proliferation of fibroblasts is dependent on its interaction with Pbx [22]. Consistent with the ability of Pbx to enhance the DNA-binding affinity of En, the co-overexpression of *PBX1b* enhanced all of the *En-1* and *En-2*-induced effects, beyond those determined for cells transduced with *En-1* or *En-2* alone (Fig. 3.5). These results, like those seen with several Hox genes in fibroblasts, suggest a genetic collaboration between *En* and *Pbx1b* in enhancing cell proliferation.

Ectopic *En-2* expression inhibits differentiation of mammary epithelial cells and renders them sensitive to 17-AAG treatment

Also restricted to HC11 cells was a noticeable morphological change upon *En-2* expression (and also with *En-1*, data not shown). 15±4% of the *En-2*-transduced heterogeneous population showed larger nuclei and reduced cytoplasm compared to the parental cells (data not shown). This morphological change is not due to the induction of a differentiation program as RT-PCR analysis shows that neither *WAP* (whey acidic protein) nor *β-casein*, differentiation markers that are rapidly induced upon hormone stimulation, are detectable in HC11 *En-2* cells (data not shown). Moreover, HC11 cells engineered to express *En-2* failed to acquire a cuboidal appearance when exposed to the lactogenic hormone cocktail (DIP; dexamethasone, insulin and prolactin, Fig. 3.5e, top

panel) but rather maintain an elongated phenotype (Fig. 3.5e, lower panel). Consistent with this observation, control HC11 cells synthesize *β-casein* transcripts in response to DIP stimulation, while *β-casein* remains undetectable in HC11 cells expressing *En-2* (Fig. 3.5f). Ectopic expression of *En-2* thus inhibits the DIP-induced differentiation program of HC11 cells *in vitro.*

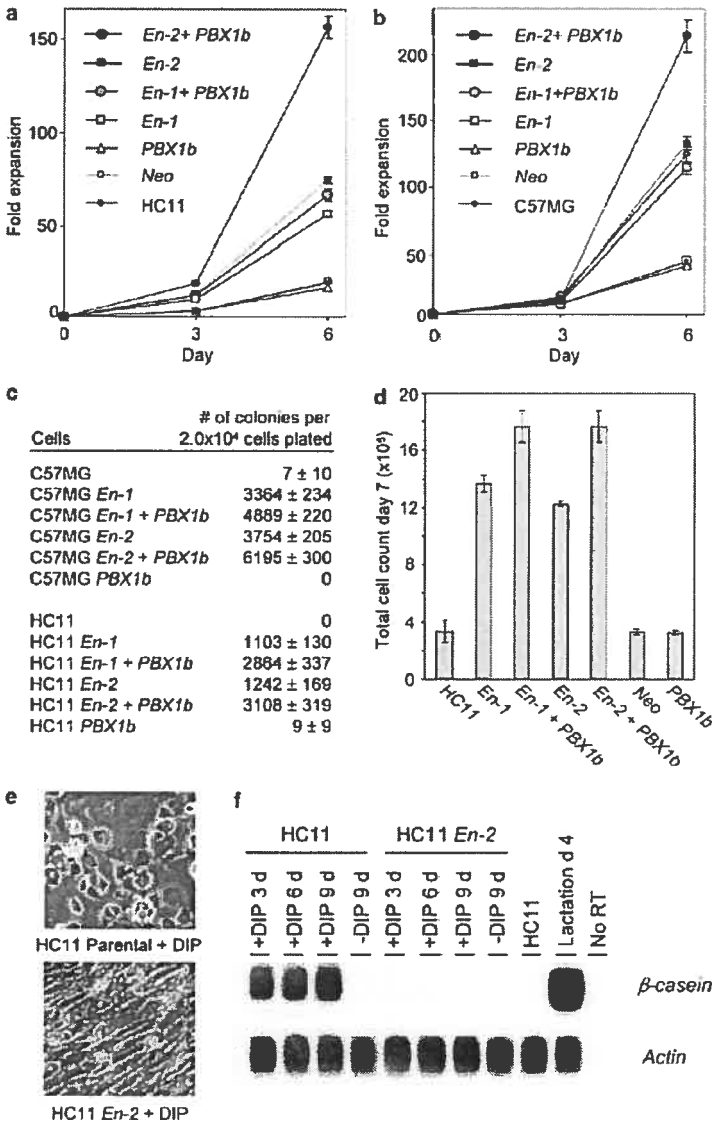

Fig. 3.5 Ectopic expression of *En-2* readily transforms HC11 and C57MG cell lines.

(a, b) Proliferation curves for HC11 and C57MG cells transduced with *En-1* and *En-2* +/- *PBX1b*. (c) Colony formation of the transduced cells after 21 days in soft agar. (d) HC11 *En-1* and *En-2* +/- *PBX1b* transduced cells show loss of cell contact inhibition. All experiments were performed immediately after selection in G418 and/or puromycin. Results in panels a-d represent the mean value ± s.e. of three separate experiments repeated in triplicate each time, with two independently infected and selected polyclonal populations. (e) HC11 parental cells stimulated with the lactogenic hormones dexamethasone (D), insulin (I) and prolactin (P) for 4 days show the characteristic large round cells undergoing differentiation which arise from a typical cuboidal epithelial-like morphology. HC11 cells expressing *En-2* maintain a more elongated fibroblastic-like morphology after the same treatment. Magnifications of both cell populations in culture were 100x. (f) RT-PCR analysis of total RNA from HC11 parental cells and HC11 cells transduced with *En-2* following 3, 6 and 9 days of treatment with DIP.

We also examined the effect of 17-AAG, an inhibitor of Hsp90 which selectively kills cancerous cells, on the growth of these mammary epithelial cell lines [23]. Proliferation of both non-transformed HC11 and C57MG cell lines was not significantly affected by 17-AAG treatment yet the proliferative advantage conferred by *En-2* expression in these two cell lines is abrogated by 17-AAG exposure (Fig. 3.6a and b). Interestingly, in the case of *En-2*-expressing C57MG cells, exposure to 17-AAG did not only revert the proliferative advantage conferred by *En-2*, but further inhibited proliferation of these cells to levels that are now below those measured in the untransduced parental cells (Fig. 3.6b). Since Hsp90 found in tumor cells has a much higher affinity for 17-AAG than the Hsp90 found in normal cells, this supports the observation that *En-2* behaves as an oncogene *in vitro* and readily transforms these mammary epithelial cells.

Transplanted *En-2*-transduced HC11 cells generate adenocarcinomas that metastasize

HC11 mammary cells introduced into surgically cleared fat pads of female hosts, under endogenous hormonal influences and within the natural microenvironment, will generate

mammary epithelial outgrowths [17]. Using this technique in an initial cohort, we observed that the majority (14/16) of mammary glands reconstituted with either *En-2* or *En-2* + *PBX1b* transduced HC11 cells developed palpable adenocarcinomas at 14 weeks while those receiving control HC11 cells (either *Neo*-transduced or parental cells) produced reconstituted glands but remained tumor-free (Fig. 3.7a).

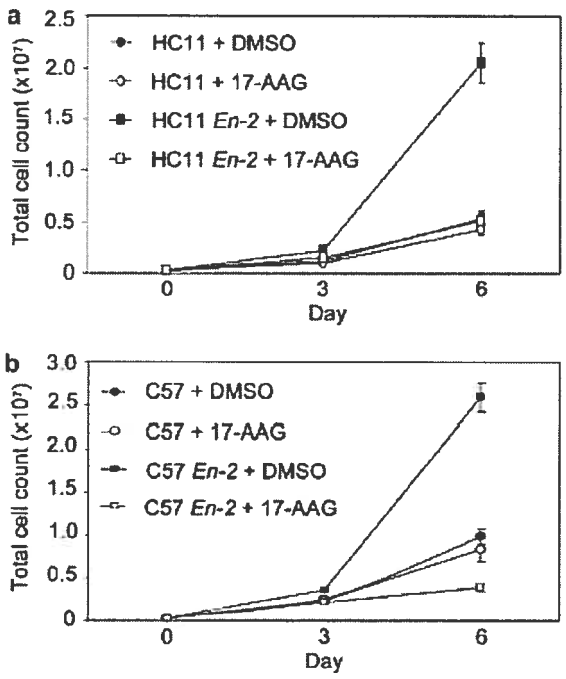

Fig. 3.6 The proliferative advantage that accompanies ectopic *En-2* expression in mammary epithelial cells is lost after exposure to 17-AAG. (A,B) Selective *in vitro* antiproliferative activity of 17-AAG in *En-2*-transduced HC11 and C57MG mouse mammary cell lines. Proliferation curves were generated from total cell counts taken 3 and 6 days following the addition of either DMSO or 5 μm 17-AAG.

To test for tumor progression in this *in vivo* tumor model a second cohort of mice were generated and sacrificed at 21 weeks post transplantation. Once again, the control groups remained tumor-free. The *En-2* and *En-2* + *PBX1b* groups developed large adenocarcinomas (32/34) with extensive fibrosis and neo-vascularisation (Fig. 3.7a and b). Macroscopic metastases were observed in multiple sites with regional nodes and

lungs being preferentially targeted by the cells initiating metastatic growth (n=6/32 mice or 19%). Metastases were also detected in the spleen and the mesenteric lymph nodes in certain cases (hollow diamonds around black diamonds indicate mice with metastases in Fig. 3.7b).

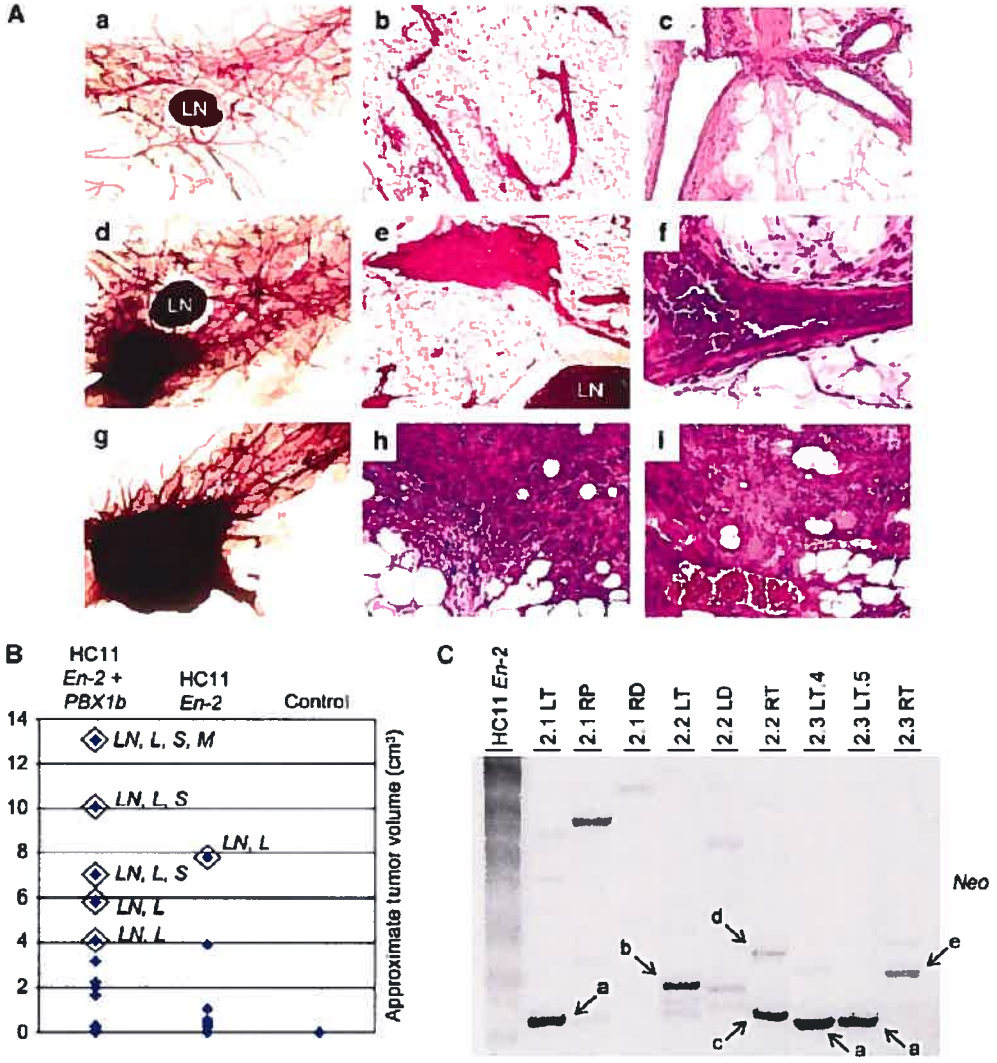

Fig. 3.7 Mammary glands reconstituted with either *En-2* or *En-2 + PBX1b* transduced HC11 cells develop adenocarcinomas. (a) Carmine Red-stained whole mounts and subsequent sections of the reconstituted inguinal mammary glands. (*a*) Mammary gland reconstituted with *Neo* transduced HC11 cells after 21 weeks (5x). (*b, c*) Subsequent sections derived from the whole mount (10x, 20x). (*d*) Mammary gland reconstituted

with *En-2* transduced HC11 cells showing one of the smaller palpable tumors after 14 weeks (5x). (*e*) H&E-stained histological section of a representative small palpable lesion in an HC11 *En-2* recipient after 14 weeks. The lesion resembles carcinoma *in situ* and is composed of hyperplastic epithelium within a thick, collagenous fibrosis (10x). (*f*) Magnification of an occluded duct seen throughout many of the recipients of *En-2*-transduced HC11 cells (40x). (*g*) Mammary gland reconstituted with *En-2* + *PBX1b* cells shows one of the larger tumors arising in the proximal region of the inguinal gland after 14 weeks (5x). (*h, i*) H&E-stained sections of representative lesions in recipients receiving HC11 cells transduced with *En-2* + *PBX1b* after 21 weeks. The tumors are large, poorly differentiated, predominantly solid nests with extensive fibrosis and surrounding angiogenesis (20x). (b) Approximate tumor volumes in the recipients after 21 weeks (volume = width^2 x length x 0.52). All the control groups (HC11 untransduced parental cells, *Neo* or *PBX1b*-transduced HC11 cells) were tumor-free. (c) Clonal analysis by Southern hybridization of DNA isolated from the resulting tumors and different regions of the reconstituted mammary gland shows the contribution of *En-2*-provirally-marked cells to three of the HC11 *En-2* recipients; 2.1, 2.2 and 2.3. The DNA is digested with BglII, which cuts only once within the provirus, and hybridization with a *Neo* probe allows the identification of different integration events and distinct clones. Exposure time was 3d at –70°C. LN, lymph node, L; lung, S: spleen, M: mesenteric lymph node, LT; left tumor, RT; right tumor, RP; right proximal, RD; right distal, LD; left distal; LT.4 left tumor in the #4 inguinal gland, LT.5; left tumor in the adjacent #5 mammary gland.

The clonal composition of the tumors and of adjacent portions of the reconstituted glands was investigated using Southern blot analysis. Proviral integration analysis showed that during the progression to tumorigenesis, typically 1-2 distinctive clones contributed to the population of the cells comprising the tumor. Such clones are distinguished by different autoradiographic signals visible at distinct sizes in different tumors (e.g., see 2 different signals in lane 7, Fig. 3.7c). From the five different tumors shown here, at least five distinct clones were identified (see clones "a" to "e" in Fig. 3.7c) indicating that there were no prominent clones in the initiating population.

Interestingly, clone "a", confirmed to be the same clone with a second digest (not shown), was found in 2 different hosts (2.1 and 2.3 in Fig. 3.7c). The presence of this clone in two different recipients could reflect a selective event which occurred *in vitro* prior to transplantation. Our inability to detect this clone in the polyclonal population from which it is derived (lane 1 Fig. 3.7c) suggests that a different selective process occurs *in vivo* versus *in vitro* and that additional genetic events were required for the development of overt mammary carcinomas in this model.

Interestingly, clones that contributed to tumor formation were different from the clones that contributed to non-hyperplastic mammary outgrowths in the more distal region of the same mammary gland (Fig. 3.7c, see 2.2 LT versus 2.2 LD). Glands transplanted with control *Neo*^{r}-transduced HC11 cells also tended to display clonal reconstitution (data not shown). Together, these results suggest that reconstitution of typical epithelial breast structures by HC11 cells depends either on the selection of a subset of "stem" cells in this population (heterogeneity), or alternatively, that these cells require adaptation to grow *in vivo*.

The clonal composition of metastases was generally identical to that of the dominant clone present in the tumor (data not shown). Importantly, metastases mostly occurred in mice that harbored large (>4 cm^3) tumors (see Fig. 3.7b). This suggests that further genetic events were needed for the tumor cells to acquire the ability to metastasize and validates *En-2* as a candidate oncogene that fosters conditions required for tumor progression *in vivo*.

siRNA-mediated suppression of *EN2* inhibits proliferation of human breast cancer cells

We next sought to determine the effect of knocking down *EN2* expression in one of the human breast cancer-derived cell lines using an RNA interference (RNAi) approach. The small interfering RNAs (siRNAs) appear to mimic intermediates in the RNAi

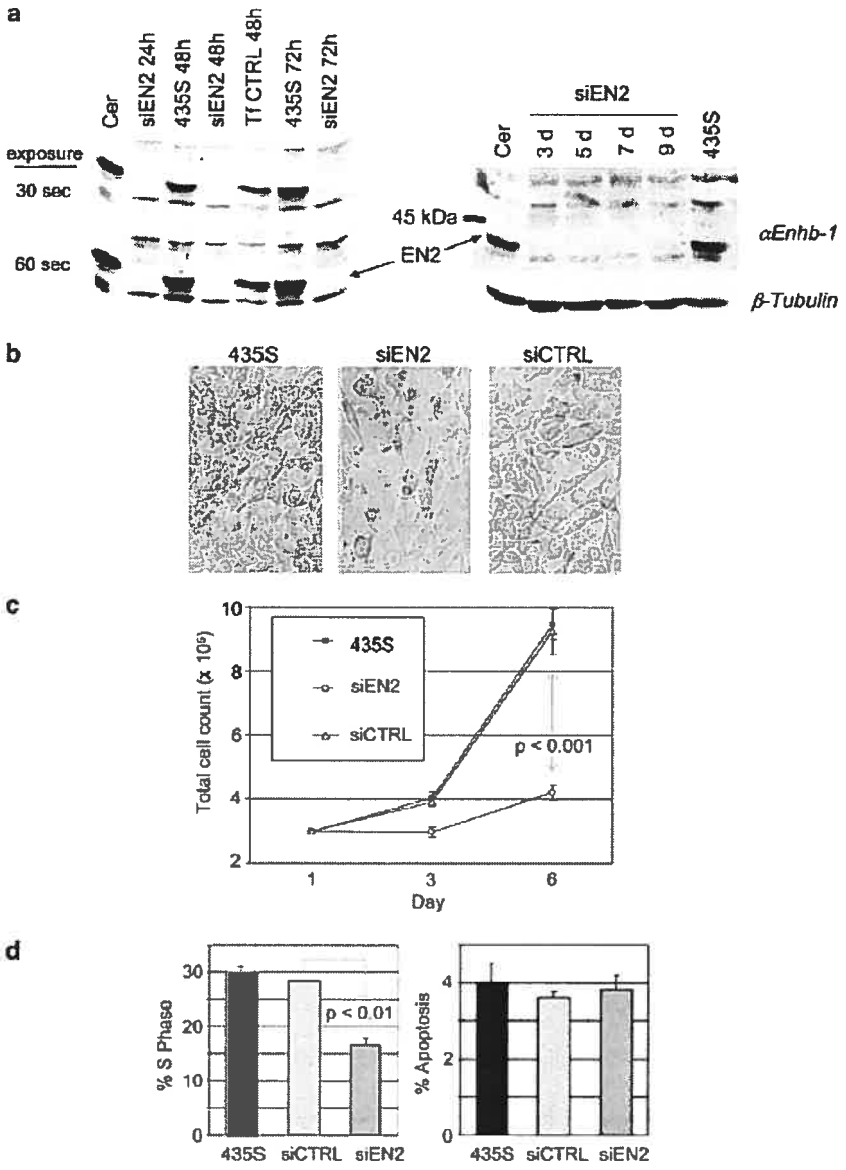

Fig. 3.8 *EN2* expression is required for proliferation of MDA-MB-435S cells. (a) Western blot analysis of the resulting transfected cells shows that siEN2 leads to specific and complete suppression of EN2 levels for up to 9 days post transfection. (b) Transfection of siEN2 leads to a change in cellular morphology as seen here 9 days post transfection. Magnifications of the cell populations in culture were 40x for 435S and 100x for both siEN2 and siCTRL. (c) Proliferation curves where total cell counts were taken at day 3 and 6 (6 and 9 days post transfection, respectively). (d) Cell cycle-DNA

content analysis and Annexin V staining performed 3 days post transfection. Cer; cerebellum, Tf CTRL; mock transfection control with 23mer ds DNA, siCTRL; scrambled control siRNA.

pathway and can silence genes in somatic cells without activating non-specific suppression by double-stranded RNA-dependent protein kinase [24]. MDA-MB-435S cells, which express high levels of *EN2* (Fig. 3.3b), were chosen as it was possible to achieve more than 90% transfection efficiency in these cells. Transient transfection of the synthetic siRNA directed against *EN2*, resulted in a reproducible and complete ablation of the protein within 48 hours which lasted for up to 12 days (Fig. 3.8a and data not shown). Microarray analysis comparing parental MDA-MB-435S cells and siEN2-transfected-MDA-MB-435S showed that siRNA exposure did not activate interferon genes such as MHC class I, oligoadenylate synthetase or enolase and that off target effects are unlikely since there was very little difference in the global transcriptome in response to siRNA treatment (data not shown).

siEN2-transfected-MDA-MB-435S cells uniquely exhibited a more flattened and retractile morphology while cells transfected with a control scrambled siRNA displayed little observable change in their transformed morphology (Fig. 3.8b). Importantly, the suppression of EN2 in these cells also resulted in a reproducible and significant decrease in their proliferation rate when compared to controls (those transfected with the scramble siRNA or mock transfected cells, Fig. 3.8c). In line with these results, cell cycle analysis revealed a 2-fold reduction in the proportion of siEN2 transfected cells in S phase when compared to controls (30% vs 16% respectively, $P<0.01$, Fig. 3.8d). This reduction in proliferation is not accompanied by an increase in apoptosis, as revealed by Annexin V staining and the lack of an increasing sub-G_1 population (right panel; Fig. 3.8d). Persistent expression of *EN2* is thus required even in this well-established breast cancer cell line.

Discussion

We have shown that *EN2* is expressed in the majority of human breast tumor-derived

cell lines and that it is ectopically expressed in ~7.3% of primary breast cancers. We have also shown that ectopic expression of *En-2,* at levels similar to those observed in primary tumors, readily transforms HC11 and C57MG cells and inhibits a differentiation program in HC11 cells that is normally induced by lactogenic hormones. *En-2* thus enhances proliferation and inhibits differentiation of mammary epithelial cells. Furthermore, our *in vivo* studies and clonal analysis of mammary adenocarcinomas occurring with *En-2*-transduced HC11 cells generated a unique model of breast cancer progression from selection of long-term repopulating cells to tumor development and to metastasis. RNA interference-mediated down-regulation of EN2 in a human breast tumor-derived cell line leads to a dramatic reduction in cell proliferation and loss of transformed morphological characteristics.

To our knowledge, *EN2* is the first candidate oncogene identified in breast cancer which is not normally expressed in breast epithelium. Southern blot analysis of genomic DNA isolated from the seven different *EN2*-positive human breast cancer cell lines studied herein failed to reveal any anomaly, suggesting that neither rearrangement nor amplification are responsible for ectopic *EN2* expression. Although the basis for overexpression of *MYC*, *Cyclin D1* and *ErbB2* is often amplification of the gene, overexpression is also observed in the absence of amplification [25]. Epigenetic modification of the *EN2* locus remains a real possibility. Very little is known about upstream regulators of *En,* in addition to Wnt, and our studies suggest that the activation of one Wnt member, Wnt-1, is not involved since *En-2* expression is not found in mouse mammary tumors induced by Wnt-1 and this oncogene is not expressed in human breast carcinomas. It will therefore be challenging to identify the exact mechanism(s) underlying the ectopic expression of *EN2*, but standard epigenetic tests will likely reveal that the locus is rendered transcriptionally active. It will be interesting to discriminate whether *EN2* is activated in preneoplastic breast lesions as both MCF 10A and MCF-12A express *EN2* and are derived from women with epithelial hyperplasia of the breast. This is in contrast with other oncogenes (e.g., *ErbB2*) which are rarely detected within early lesions or benign breast disease [26] and may indicate that ectopic *EN2* expression occurs early in breast cancer.

The reasons for the high incidence of *EN2* expression in breast cancer cell lines (as opposed to primary tumors) are not clear. Perhaps a selection process (proliferation, independence of stroma, etc.) favors the emergence of cancer cell lines which express this gene. As the notion of breast cancer stem cells is emerging [27], it will become interesting to investigate whether *EN2* determines the proliferative activity of these tumor stem cells, as we recently showed for *Bmi-1* in leukemic stem cells [28].

Of interest, *EN2* was also detected in SAGE libraries derived from human brain glioblastoma, colon and ovarian carcinoma (Library numbers NCI CGAP Bm23, NCI CGAP Co16 and CL ES2-1, respectively, http://www.ncbi.nlm.nih.gov/UniGene). It will be essential to analyze whether *EN2* is also ectopically expressed in other epithelial tumors as the observations described herein may be generalized to other carcinomas. Importantly, it will be critical to determine an even more accurate frequency of *EN2* ectopic expression in breast cancers using a larger number of specimens and to assess whether its expression correlates predominantly to infiltrating ductal carcinomas. Similar to the current rationale that herceptin and 17-AAG, in combination, will increase the taxol response even further in *HER-2* overexpressing breast and prostate cancers, the combination of 17-AAG and siRNA against EN2 may constitute a strong assault against *EN2*-positive breast tumors. Although siRNA directed against EN2 offers great specificity in cancerous lesions detected early, as we have shown that *EN2* is ectopically expressed in breast cancer, the addition of 17-AAG may also affect those tumors that have evolved to evade the toxic effects of single molecularly targeted agents. This approach could have great clinical benefit as if it does not kill the cancer cells, it might leave them sufficiently debilitated and more sensitive to chemotherapy and radiotherapy [23].

Materials and Methods

Transgenic mice

The generation of MMTV-Wnt-1 transgenic mice has been reported previously [16] and were purchased from The Jackson Laboratory (FVB/N hybrid background, stock # 002934). MMTV-Wnt-1 mice were genotyped by Southern blot analysis of Bam HI-digested tail DNA using a 904-bp Kpn I-Sph I fragment of SV40 poly(A) DNA. BALB/c mice were acquired from Harlan Labs (Indianapolis, IN). All animals were maintained and bred in ventilated microisolator cages, provided with sterilized food and acidified water in the specific pathogen-free (SPF) animal facility of the Clinical Research Institute of Montreal (IRCM).

Cell lines

The HC11 mammary epithelial cell line is a clonal derivative of the COMMA-1D cell line, derived from mammary tissue of a mid-pregnant BALB/c female [29,30]. The C57MG cell line was derived from glands of a 23-week-old retired C57BL/6 breeder [31]. MCF7, MDA-MB-231, SK-BR-3, MDA-MB-468, MDA-MB-435S, MDA-MB-436, and BT-20 human epithelial cell lines were derived from breast adenocarcinomas. T-47D and BT-474 human epithelial cell lines were isolated from ductal carcinomas. MCF 10A and MCF-12A human epithelial cell lines were derived from fibrocystic breast tissue and both lines have been reported to form colonies in soft agar. The HBL 100 human epithelial cell line was originally derived from breast milk, yet also forms colonies in soft agar. C57MG, HC11, T-47D, HBL 100, and MCF7 lines were grown in RPMI 1640 medium supplemented with 10% fetal calf serum, 10 ng/ml of epidermal growth factor and 10 µg/ml of insulin. MDA-MB-435S, MDA-MB-436 and MDA-MB-468 lines were grown in Lebowitz-L15 medium supplemented with 10% fetal calf serum. MCF 10A and MCF-12A lines were grown in F12 HAM:DMEM (1:1) medium supplemented with 5% fetal calf serum. The BT-20 line was grown in αMEM medium supplemented with 10% fetal calf serum.

Generation of recombinant retroviruses and infection of mammary cell lines

The entire coding regions of the mouse *En-1* (nucleotides 274-1578; Accession no. L12703, #552) and *En-2* (nucleotides 1-1315; Accession no. L12705, #547; the mouse

En-2 protein shares 90% amino acid sequence identity with the human EN2 protein) cDNAs were introduced into the Hpa I and Hpa I-Bgl II sites, respectively, downstream of the retroviral promoter contained within the 5' long terminal repeat (LTR), of the MSCVneoEB retroviral vector which confers G418 resistance under the control of the phosphoglycerate kinase (PGK) promoter. The MSCV-human *PBX1b*-PGK-PAC retroviral vector (which confers puromycin resistance) was described previously [22]. High-titer helper-free recombinant retroviruses were produced from BOSC-23 viral packaging cells and tested as previously reported [22]. HC11 and C57MG cell lines were infected by exposure to filtered (0.2 μm, low-protein binding filter, Millipore, Bradford, MA) viral supernatant in the presence of 6 μg/ml polybrene (Sigma). Transduced cells were selected and maintained in 220 and 260 μg/ml of G418 for HC11 and C57MG lines, respectively, or 2.5 μg/ml puromycin, or both drugs concurrently, as appropriate for selection of virus encoded selectable markers.

Southern, Northern and Western Blot Analysis

To assess proviral integration, Southern hybridization analyses were performed as previously described [32]. 10 μg of genomic DNA was digested with Kpn I or Nhe I which cleaves in both flanking LTRs to release the provirus. Membranes were hybridized with *Neo-* or *Puro-* specific probes labeled with ^{32}P-dCTP by random primer extension as described [33]. Following autoradiography, blots were stripped and hybridized using a probe specific to *HoxA9* (1.1 kb Hind III fragment) to assess loading. For Northern blot analysis, 10 μg total RNA isolated with TRIzol (GIBCO) was separated on a 1% formaldehyde-agarose gel as described [34] and hybridized with a 186-bp Bgl II *En-1* cDNA (#552) probe, a 254-bp Bgl II-Sst I *En-2* cDNA (#530) probe and a 1.6-kb Bgl II-Eco RI *PBX1b* cDNA (#448) probe. After autoradiography, the blots were stripped and rehybridized with an oligonucleotide complementary to 18S rRNA [34]. For western blot analysis, total and nuclear extracts were prepared as reported previously [35]. 100 μg total and 40 μg nuclear aliquots of protein were separated by SDS-PAGE as described [36]. En proteins were detected with α*Enhb-1* antisera (which detects both 41 kDa mouse En-1 and human EN1, and 55 kDa mouse En-2 and human EN2 proteins) as described [37]. PBX1b proteins were detected with an anti-PBX1 polyclonal antibody (P-20; cat# sc-

889; Santa Cruz Biotechnology Inc, Santa Cruz, Calif.). As a control for loading, all membranes were stripped and hybridized with αPTP1D (protein-tyrosine phosphatase 1D; P54420; BD PharMingen, Mississauga, Canada).

cDNA generation, amplification and analysis

Total RNA was isolated from adult mouse cerebellum, mouse mammary glands, cell lines, frozen human primary breast tumors, adjacent normal breast samples and frozen human reduction mammoplasty tissue using TRIzol. Reverse transcription and amplification of 0.1 μg of the resulting total RNA were performed as described previously [38]. Single-copy probes corresponded to a 436-bp fragment of the mouse *β*-casein cDNA (nt 4871-5307; Accession no. M26940 X13484, #1051) and *Actin*, isolated as described in [39]. The amplification of *Actin* was used as a control for both quality and quantity of templates in each sample. To demonstrate that the amplification was solely from cDNA and not from DNA contamination, a control which contained RNA but no reverse-transcriptase (No RT) was included in each experiment.

Immunohistochemistry

Frozen sections were cut at 5 μm and were subsequently fixed briefly in paraformaldehyde. Immunohistochemistry was performed using a three-step streptavadin-biotin peroxidase method and antigen retrieval was carried out by microwave heating in citrate buffer. Primary antibody rabbit anti-mouse polyclonal *αEnhb-1* was used at a final dilution of 1/500. Biotinylated goat anti-rabbit IgG secondary antibody (Vector Laboratories, Burlingame, CA) was used at a final dilution of 1/150, and revealed using Streptavidin-Horseradish Peroxidase (NEL 750, NEN) at 1/1000. Slides were counter-stained with Methyl Green.

Soft agar colony formation, contact inhibition and proliferation assays

For proliferation assays, the selected HC11 and C57MG polyclonal transduced cell populations were trypsinized and replated at $3x10^5$ and 5000 cells per 10 cm^2 dish in

RPMI 1640 with 5% FBS, for 3 and 6 days of growth respectively, and subsequently starved, serum stimulated and counted as previously described [22]. For colony assays, the cells were grown for 3 days in the absence of antibiotic selection and then $2x10^4$ cells were plated in RPMI 1640 medium supplemented with 10% fetal calf serum, 10 ng/ml of epidermal growth factor and 10 μg/ml of insulin containing 0.3% Agar Noble into 35 mm^2 Petri dishes containing a layer of solidified 0.6% agar. Colonies were scored 21 days after being plated using a surface area that corresponded to 1/8 of the $35mm^2$ dish using an inverted microscope. For the contact inhibition assay, selected HC11 polyclonal transduced cell populations were trypsinized and replated at near confluence in 10 cm^2 dishes in RPMI 1640 with 10% FBS. Total cell counts were taken 7 days later when dense foci were readily visible in the cultures that no longer displayed contact inhibition and continued to grow.

Lactogenic hormone stimulation of HC11 mammary epithelial cells

HC11 cells and HC11 cells expressing *En-2* were grown to confluency in 10 cm^2 dishes and maintained for 3 days in normal media. Confluent cultures were washed and incubated for 18 hr in serum-free media (RPMI 1640 containing 1 mg/ml fetuin and 10 μg/ml transferrin) followed by 3, 6 and 9 days of treatment with induction medium (RPMI 1640 containing 10^{-6} M dexamethasone, 5 μg/ml insulin, and 5 μg/ml ovine prolactin /luteotropic hormone; Sigma) as described [40]. Parallel unstimulated controls were subjected to the same regimen but were kept in RPMI 1640 with 5% FCS after serum-free starvation. The morphological changes in HC11 cells expressing *En-1*, *En-1+ PBX1b*, *En-2* and *En-2 + PBX1b* were scored on cytospin preparations containing 150,000 cells (n=4 slides for each population).

Transplantation of HC11 Transduced Cells into Syngeneic Hosts

Selected polyclonal populations of HC11 cells transduced with *En-2*, *En-2 + PBX1b*, *PBX1b*, *Neo* and untransduced HC11 cells were collected from 10 cm^2 dishes and resuspended in normal growth medium at a final concentration of 5 x 10^5 cells/10 μl. Using a beveled syringe, the cells were injected into the cleared fat pads of female

BALB/c syngeneic hosts, just above the lymph node, in a volume of 10 µl. The surgical procedures for clearing the endogenous mammary epithelium from the #4 (inguinal) fat pads of 3-week-old female mice have been described [41]. In each case, whole-mount preparations of the excised host mammary epithelium were generated to verify complete removal. Two cohorts of mice were sacrificed at 14 and 21 weeks post transplantation, respectively, and the glands were surgically removed. Whole-mount preparations and subsequent sections were produced from the reconstituted glands in the first cohort. In the second cohort, the majority of the resulting tumors, different portions of the reconstituted glands, and potential sites of metastases were either fixed in paraffin and subsequently sectioned, or used for DNA extraction.

Whole Mounts and Tumor Histology

Inguinal mammary glands were resected and flatten fixed in Carnoy's fixative, defatted in ethanol and acetone, rehydrated and stained in Carmine Red. The mammary whole mounts were reprocessed for paraffin embedding and 5 µm sections were prepared. Tumors and potential sites of metastases (brain, lung, femur, spleen, lymph nodes, and the #5 mammary gland) were fixed O/N in 4% PFA, embedded in paraffin, sectioned at 5 µm and stained by H&E.

RNA Interference Studies

The 21-nt human *EN2* target sequence used to design the synthetic siRNA was 5'-AAC TTC TTC ATC GAC AAC ATC-3'. The selected sequence was subjected to a BLAST search against the human genome sequence to ensure that only *EN2* would be targeted. The 21-nt sequence constituting the control scrambled siRNA (siCTRL) was 5'- AA GCG CGC TTT GTA GGA TTC G -3'. Synthetic siRNA oligonucleotides were purchased from Dharmacon (Lafayette, CO). MDA-MB-435S cells were regularly passaged to ensure exponential growth and were passaged the day before transfection. Subconfluent MDA-MB-435S cells were transfected with 150 nM siRNA / 6 cm^2 dish and fresh media was provided 36 h after (details of the transfection procedure are

available upon request to S.M.). Whole-cell extracts were prepared 3, 5, 7, 9 and 12 days after transfection, separated on a 10% SDS-PAGE and immunoblotted to reveal EN2 protein. The same membrane was immunoblotted with antibody against β-Tubulin as a control for loading. Cells were harvested for flow cytometry and plated for proliferation assays 3 days after transfection in at least three independent experiments.

Cell Cycle Analysis and Annexin V Staining

MDA-MB-435S cells were trypsinized three days post-transfection, washed twice with PBS, and incubated for 30 min on ice in hypotonic DNA staining solution (0.1% Sodium citrate, 0.3% NP-40, 0.02 mg/mL RNase A, 50 ug/mL Propidium Iodide). Stained nuclei (10 000/sample) were analysed by flow cytometry. Parallel 6 cm^2 dish were trypsinized, washed twice with PBS, and incubated 15 min on ice in Annexin V binding buffer (10 mM Hepes pH 7.4, 150 mM NaCl, 5 mM KCl, 1 mM $MgCl_2$, 1.8 mM $CaCl_2$, 2.5 µg/ml Annexin V-FITC; 556419 BD PharMingen, 50 µg/ml Propidium Iodide). Stained cells (10 000/sample) were analyzed by flow cytometry to detect phosphatidyl serine exposure and damaged cell membranes.

Acknowledgements

We thank Dr. Alexandra Joyner (NYU, NYC) for *En-1* and *En-2* cDNAs and α*Enhb-1* antisera; Dr. Morag Park (McGill University, Montréal) for the BT-474, MDA-MB-231, SK-BR-3, MDA-MB-468, MDA-MB-435S, MCF 10A, MCF-12A, MDA-MB-436, BT-20 cell lines; Dr. Morag Park and Dr. Michel Trudel (Montréal General Hospital, Montréal) for primary human breast tumor and normal adjacent tissue samples; Dr. Shiu (University of Manitoba, Winnipeg) for the MCF7 and HBL 100 cell lines; Dr. Paul Jolicoeur (IRCM) for the HC11 cell line; Dr. Anthony Brown (Cornell University, NYC) for the C57MG cell line; Dr. Robert Oshima (Burnham Institute, La Jolla) for human *Keratin 18* cDNA; Christian Charbonneau for assistance in photography, and Annie Vallée and Martin Demers for histological sectioning services. This work was supported by funds from the NCIC (15064) to G.S., and the NCIC to S.M. N.M. is a

recipient of a CIHR research award, G.S. is a scholar of the Leukemia and Lymphoma Society of USA, and S.M. is an Investigator of the CIHR.

References

1. Baselga,J. & Norton,L. Focus on breast cancer. *Cancer Cell.* **1**, 319-322 (2002).

2. Tognon,C. et al. Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell.* **2**, 367-376 (2002).

3. Liao,D.J. & Dickson,R.B. c-Myc in breast cancer. *Endocr.Relat Cancer.* **7**, 143-164 (2000).

4. Hutchinson,J.N. & Muller,W.J. Transgenic mouse models of human breast cancer. *Oncogene.* **19**, 6130-6137 (2000).

5. Bieche,I. et al. Quantitation of MYC gene expression in sporadic breast tumors with a real-time reverse transcription-PCR assay. *Cancer Research.* **59**, 2759-2765 (1999).

6. Lizard-Nacol,S. et al. Benign breast disease: absence of genetic alterations at several loci implicated in breast cancer malignancy. *Cancer Res.* **55**, 4416-4419 (1995).

7. Nusse,R. & Varmus,H.E. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell.* **31**, 99-109 (1982).

8. Huguet,E.L. et al. Differential expression of human Wnt genes 2, 3, 4, and 7B in human breast cell lines and normal and disease states of human breast tissue. *Cancer Res.* **54**, 2615-2621 (1994).

9. Cadigan,K.M. & Nusse,R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11**, 3286-3305 (1997).

10. He,T.C. et al. Identification of c-MYC as a target of the APC pathway. *Science.* **281**, 1509-1512 (1998).

11. Tetsu,O. & McCormick,F. Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature.* **398**, 422-426 (1999).

12. McMahon,A.P., Joyner,A.L., Bradley,A. & McMahon,J.A. The midbrain-hindbrain phenotype of Wnt-1-/Wnt-1- mice results from stepwise deletion of engrailed-expressing cells by 9.5 days postcoitum. *Cell.* **69**, 581-595 (1992).

13. Davis,C.A. & Joyner,A.L. Expression patterns of the homeo box-containing genes En-1 and En-2 and the proto-oncogene int-1 diverge during mouse development. *Genes Dev.* **2**, 1736-1744 (1988).

14. Joyner,A.L. et al. Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the En-2 homeobox. *Science.* **251**, 1239-1243 (1991).

15. Lewis,M.T. Homeobox genes in mammary gland development and neoplasia. *Breast Cancer Res.* **2**, 158-169 (2000).

16. Tsukamoto,A.S. et al. Expression of the int-1 gene in transgenic mice is associated with mammary gland hyperplasia and adenocarcinomas in male and female mice. *Cell.* **55**, 619-625 (1988).

17. Humphreys,R.C. & Rosen,J.M. Stably transfected HC11 cells provide an in vitro and in vivo model system for studying Wnt gene function. *Cell Growth Differ.* **8**, 839-849 (1997).

18. Hynes,N.E. & Stern,D.F. The biology of erbB-2/neu/HER-2 and its role in cancer. *Biochim.Biophys.Acta.* **1198**, 165-184 (1994).

19. Brown,A.M., Wildin,R.S., Prendergast,T.J. & Varmus,H.E. A retrovirus vector expressing the putative mammary oncogene int-1 causes partial transformation of a mammary epithelial cell line. *Cell.* **46**, 1001-1009 (1986).

20. Peltenburg,L.T. & Murre,C. Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J.* **15**, 3385-3393 (1996).

21. Peltenburg,L.T. & Murre,C. Specific residues in the Pbx homeodomain differentially modulate the DNA-binding activity of Hox and Engrailed proteins. *Development.* **124**, 1089-1098 (1997).

22. Krosl,J. et al. Cellular proliferation and transformation induced by HOXB4 and HOXB3 proteins involves cooperation with PBX1. *Oncogene.* **16**, 3403-3412 (1998).

23. Neckers,L. & Lee,Y.S. Cancer: the rules of attraction. *Nature.* **425**, 357-359 (2003).

24. Elbashir,S.M. et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature.* **411**, 494-498 (2001).

25. Ross,J.S. et al. Breast cancer biomarkers and molecular medicine. *Expert.Rev.Mol.Diagn.* **3**, 573-585 (2003).

26. Ingvarsson,S. Molecular genetics of breast cancer progression. *Semin.Cancer Biol.* **9**, 277-288 (1999).

27. Al Hajj,M. et al. From the Cover: Prospective identification of tumorigenic breast cancer cells. *Proc.Natl.Acad.Sci.U.S.A.* **100**, 3983-3988 (2003).

28. Lessard,J. & Sauvageau,G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature*. **423**, 255-260 (2003).

29. Danielson,K.G. et al. Epithelial mouse mammary cell line exhibiting normal morphogenesis in vivo and functional differentiation in vitro. *Proc.Natl.Acad.Sci.U.S.A.* **81**, 3756-3760 (1984).

30. Ball,R.K. et al. Prolactin regulation of beta-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. *EMBO J.* **7**, 2089-2095 (1988).

31. Vaidya,A.B., Lasfargues,E.Y., Sheffield,J.B. & Coutinho,W.G. Murine mammary tumor virus (MuMTV) infection of an epithelial cell line established from C57BL/6 mouse mammary glands. *Virology*. **90**, 12-22 (1978).

32. Pawliuk,R., Kay,R., Lansdorp,P. & Humphries,R.K. Selection of retrovirally transduced hematopoietic cells using CD24 as a marker of gene transfer. *Blood.* **84**, 2868-2877 (1994).

33. Lawrence,H.J. et al. Stage- and lineage-specific expression of the HOXA10 homeobox gene in normal and leukemic hematopoietic cells. *Exp.Hematol.* **23**, 1160-1166 (1995).

34. Kroon,E. et al. Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *EMBO J.* **17**, 3714-3725 (1998).

35. Thorsteinsdottir,U. et al. The oncoprotein E2A-Pbx1a collaborates with Hoxa9 to acutely transform primary bone marrow cells. *Mol.Cell Biol.* **19**, 6355-6366 (1999).

36. Thorsteinsdottir,U. et al. The oncoprotein E2A-Pbx1a collaborates with Hoxa9 to acutely transform primary bone marrow cells. *Mol.Cell Biol.* **19**, 6355-6366 (1999).

37. Davis,C.A., Holmyard,D.P., Millen,K.J. & Joyner,A.L. Examining pattern formation in mouse, chicken and frog embryos with an En-specific antiserum. *Development*. **111**, 287-298 (1991).

38. Sauvageau,G. et al. Differential expression of homeobox genes in functionally distinct CD34+ subpopulations of human bone marrow cells. *Proc.Natl.Acad.Sci.U.S.A.* **91**, 12223-12227 (1994).

39. Sauvageau,G. et al. Differential expression of homeobox genes in functionally distinct CD34+ subpopulations of human bone marrow cells. *Proc.Natl.Acad.Sci.U.S.A.* **91**, 12223-12227 (1994).

40. Cella,N., Groner,B. & Hynes,N.E. Characterization of Stat5a and Stat5b homodimers and heterodimers and their association with the glucocortiocoid receptor in mammary cells. *Mol.Cell Biol.* **18**, 1783-1792 (1998).

41. DeOme,K.B., Faulkin,L.J., Bern,H.A. & Blair,P.B. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *J.Natl.Cancer Inst.* **78**, 515-525 (1959).

CHAPITRE 4

Identification of Putative *EN2* Transcriptional Targets by Microarray Analysis

Nicole Martin, Richard Le Blanc & Guy Sauvageau

(Manuscript in preparation)

Chapter 4 summarizes the experimental data obtained thus far in identifying putative *EN2* transcriptional targets in human breast cancer cell lines by microarray analysis. Dr. Richard Le Blanc, a Clinician Scientist affiliated with the laboratory at the time, conducted the statistical analysis of the microarray data and generated Figure 4.2, Table 4.3 and the Materials and Methods section entitled 'Expression Profiling Data Analysis'.

Introduction

Evidence has accumulated that supports the notion that homeobox genes, when dysregulated, are involved in tumorigenesis, yet the precise mechanism(s) involved are largely unknown. The molecular mechanism(s) through which *EN2* mediates breast tumorigenesis have not been elucidated and would require the identification of downstream target genes. Although *Drosophila en* has been implicated in the regulation of several pathways, few target genes have been shown to be under its direct regulatory control. Methodologies used for the isolation of target genes and for the analysis of putative targets will be beneficial in establishing the genetic pathways controlled by *EN2* within breast tissue.

Transient transfection of the synthetic siRNA directed against *EN2* resulted in a reproducible and complete reduction of EN2 protein levels in the human breast cancer-derived cell line MDA-MB-435S [1]. The specificity of siEN2 was confirmed using a control-scrambled siRNA that failed to suppress *EN2* expression. Not only was this approach more successful than anticipated, but the complete knockdown mediated by siEN2 was maintained for up to 12 days post transfection. MDA-MB-435S cells transfected with siEN2 quickly vary to exhibit a more cuboidal flattened morphology, reminiscent of non-transformed cells, while the control cells displayed little or no observable change in their transformed morphology. The suppression of EN2 in MDA-MB-435S cells also resulted in a reproducible and significant decrease in proliferation rates when compared to controls. Importantly, this reduction in proliferation was not accompanied by an increase in apoptosis. With the overt phenotypical changes seen upon suppressing EN2 protein levels in MDA-MB-435S cells, it is likely that there are key expression profile changes that take place.

We reasoned that comparing the original MDA-MB-435S cancer cell line, which expresses high levels of EN2, and the siEN2-transfected cells where EN2 has been completely knocked-down, serves as an invaluable source to identify key targets that are either up- or downregulated in response to EN2 suppression. The recent sequencing of the human genome provides a foundation for studying global patterns in gene

expression. The simultaneous measurement of the relative abundance of thousands of mRNAs allows characterization of the cellular transcriptome under various experimental conditions. Oligonucleotide arrays have been used extensively for mRNA expression analysis and the exemplar of this technology is the Affymetrix Gene Chip.

Results

Employing siRNA to elucidate potential downstream targets of EN2

Comparing the gene expression profile of the original MDA-MB-435S cancer cell line, which expresses high levels of *EN2*, with the transcriptome of the same MDA-MB-435S cell line, transfected with siEN2 to knock-down EN2 protein levels, should serve as an invaluable source to identify key targets that are either up- or downregulated in response to EN2 suppression. To further enhance this experimental design, the reciprocal approach of ectopically expressing *EN2* in T-47D cells, another human breast adenocarcinoma cell line, was carried out. This should allow one to better discern transcriptional programs that are truly relevant to breast cancer rather than those that are unique to one specific cell line or the consequence of one specific manipulation (i.e. transfection of siRNA vs retroviral transduction). RT-PCR analysis of the resulting RNA confirms that the expected experimental system was indeed attained whereby EN2 is completely suppressed in MDA-MB-435S cells and *EN2* is ectopically expressed in T-47D cells (Fig 4.1). Moreover, the level of ectopic *EN2* expression achieved in T-47D cells was comparable to that documented in parental MDA-MB-435S cells, so we could presumably rule out dosage responsive targets found only at non-physiological elevated levels.

Although the Affymetrix HG-U133A array contains approx 22 000 probesets, only a small proportion of transcripts showed a significant change in expression after normalization. The majority of these transcripts were upregulated, suggesting that although EN2 is capable of acting both as a transcriptional activator and repressor, inducing gene expression seems to be more prevalent in breast cancer cells (Table 4.1). Many genes that were significantly differentially expressed in one cell line were not

significantly differentially expressed in the second cell line or did not even respect the same trend of either upregulation or downregulation in response to EN2 (Table 4.1). Employing two breast cancer cell lines not only affords an increased chance of finding bona fide targets that are common to both cell lines while excluding targets that are specific to one cell line, but also points out how expression studies which use one cell line or tissue system will give rise to many context-specific targets.

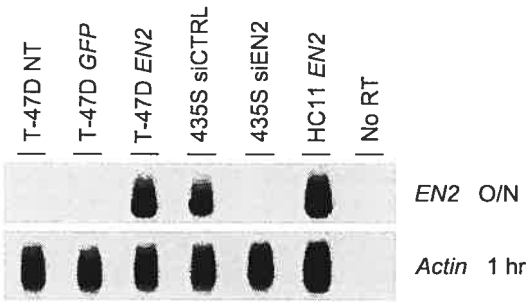

Fig. 4.1 RT-PCR analysis of the same total RNA which was used in the microarray analysis. *EN2* is successfully ectopically expressed in the T-47D human breast adeonocarcinoma cell line 3 days post infection (T-47D *EN2* lane) and completely suppressed in the MDA-MB-435S human breast adenocarcinoma cell line 3 days post transfection (435S siEN2 lane). NT, non transfected; GFP, infected with Enhanced Green Fluorescent Protein; siCTRL, transfected with control scrambled siRNA; No RT, no reverse transcriptase.

Importantly, siRNA exposure did not activate interferon genes such as MHC class I, oligoadenylate synthetase or enolase in MDA-MB-435S cells and the ectopic expression of *EN2* in T-47D cells did not activate alternate cell-specific transcriptional programs. The first 50 common ranked differentially expressed genes that were identified are listed with their corresponding ID identifiers (Fig 4.2). A large proportion of the genes found to be commonly upregulated (or induced further) in the presence of *EN2* in both cell lines have chaperone and/or heat shock activity (Table 4.2). Gene Ontology (GO) Mining Tool provided by the NetAffyx Analysis Center predicted that 1.01 genes would

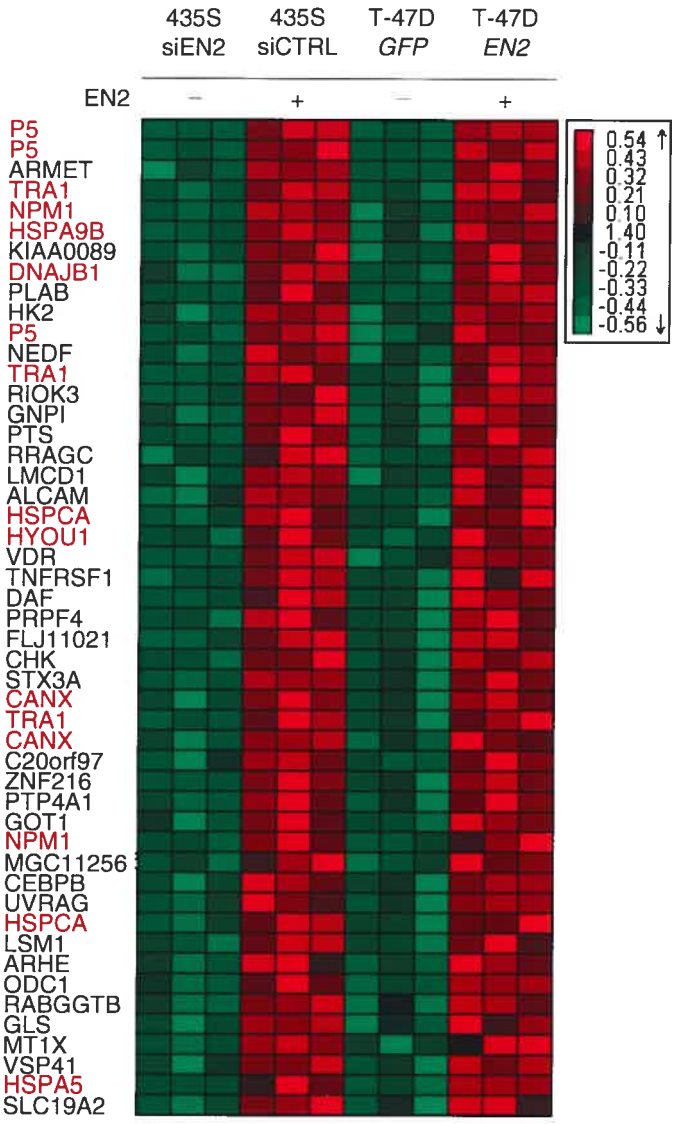

Fig. 4.2 Heat map cluster diagram of the first 50 top-ranked differentially expressed filtered putative *EN2* transcriptional targets whose expression correlates positively with the presence of EN2. Each column represents a single microarray experiment, and each row represents a single gene, where the corresponding ID identifier is listed beside each row. The color of each square in this image represents the measured expression ratio of each gene in question. The color saturation is also directly proportional to the magnitude of the measured gene expression ratio where bright red squares represent the highest R/G ratio, bright green squares represent the lowest R/G ratio of gene expression in the experimental samples and black squares represent a ratio of approximately 1.

be associated with heat shock protein activity and 4.25 genes with chaperone activity among the first 477 probesets which were significant at the nominal 0.001 level of the univariate two-sample t-test (Table 4.3) [2,3]. The actual numbers exceeded these predictions as 7 genes with heat shock protein activity and 22 genes with chaperone activity were observed in this subset, representing an observed/expected ratio of 6.92 and 5.17, respectively, over the predicted values (Table 4.3). These ratios are likely underestimated as *P5*, the top-ranked upregulated gene in our analysis (with 3 probesets amongst the first 13 top-ranking probesets), has only recently been ascribed chaperone activity [4]. Genes such as *Hsp90α*, *Hsp90β*, *Calnexin*, *DnaJ (Hsp40)*, *Nucleophosmin*, *hP5*, *Hsp70* and *Tumor rejection antigen (gp96)* all have chaperone activity or associate with chaperone proteins. Of particular interest was the dramatic induction of *Hsp90*, a molecular chaperone which has been implicated in the survival of cancer cells [5].

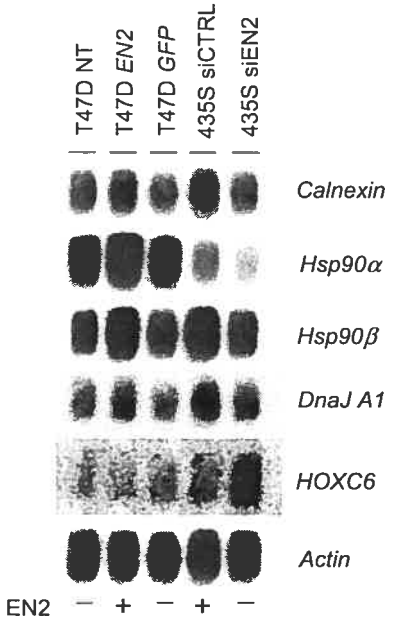

Fig. 4.3 RT-PCR analysis of the five most differentially expressed transcriptional targets identified by microarray analysis. The above induction trends for *Hsp90α*, *Hsp90β*, *Calnexin* and *DnaJ A1* and the repression trend for *HOXC6* reflect the induction and repression found in the microarray data.

To validate some of the targets acquired by our microarray analysis, RT-PCR analysis was performed on the remaining RNA that was prepared for microarray hybridization. Five of the most differentially expressed transcriptional targets identified by microarray analysis were confirmed (Fig 4.3). The same induction trend was seen for *Hsp90α*, *Hsp90β*, *Calnexin* and *DnaJ A1* while documentation of the same repression trend as that seen in the microarray data was found for *HOXC6* (Fig 4.3). RT-PCR results also verified that these five genes are regulated by *EN2* expression in both T-47D and MDA-MB-435S human breast cancer cell lines.

The upregulation of genes with chaperone and/or heat shock activity in response to *EN2* expression is not a non-specific effect as the observed upregulation is only seen in the presence of EN2 and is not induced by the stress of either siRNA or retroviral manipulation alone in the control cells. The exact mechanism by which *EN2* expression leads to a further induction of these genes and whether they are direct targets remains to be determined, but this initial observation is potentially interesting and fitting as molecular chaperones have been implicated in the survival of cancer cells and one chaperone in particular, Hsp90α, has recently been shown to function in tumor invasion [6,7,8]. Ongoing clinical trials unexpectedly found that 17-AAG, an inhibitor of Hsp90, selectively kills cancerous cells [9]. This selectivity occurs because Hsp90 found in tumor cells has a much higher affinity for 17-AAG and higher ATPase activity, required for its chaperone function, than the Hsp90 found in normal cells [10]. 17-AAG also inhibits the proliferation of human cancer cells and has shown antitumor activity in several xenograft models, including breast [11].

We previously examined the effect of 17-AAG on the growth of mammary epithelial cell lines engineered to ectopically express *EN2*. The proliferation of both non-transformed HC11 and C57MG cell lines was not significantly affected by 17-AAG treatment yet the proliferation advantage conferred by *EN2* expression in these two cell lines is abrogated by 17-AAG exposure [1]. Interestingly, in the case of *EN2*-expressing C57MG cells, exposure to 17-AAG did not only revert the proliferative advantage conferred by the oncogene, but further inhibited proliferation of these cells to levels that

are now below those measured in the untransduced parental cells [1]. Since Hsp90 found in tumor cells has a much higher affinity for 17-AAG than the Hsp90 found in normal cells, this supports the observation that *EN2* behaves as an oncogene *in vitro* and readily transforms these mammary epithelial cells. In addition, the proliferation of both T-47D and MDA-MB-435S cells expressing *EN2* was significantly further reduced in comparison to the decrease in proliferation seen in both cell lines when *EN2* is not expressed (data not shown). This observation likely reflects the increased chaperone protein expression, including Hsp90, in the two cell lines where *EN2* is expressed, which renders the cells more sensitive to 17-AAG.

Discussion

Expression profiling employing two breast tumor cell lines where *EN2* was up and down regulated highlighted a large group of chaperone and heat shock proteins, including Hsp90, whose expression are coordinately upregulated by EN2. The large number of differentially expressed genes that were found to be specific to each cell line also highlights the caveat of studies employing only one cell line or tissue in microarray analysis. The exact molecular mechanism(s) by which *EN2* fosters tumorigenicity still need to be elucidated, but this is the very first glimpse into potential mammalian targets that may convey the transcriptional signal from EN2 and identification of putative genes regulated by EN2 was a pivotal next step to providing some insight into how EN2 may contribute to the pathogenesis of breast cancer.

Expression profiling technologies offered by microarrays facilitated the identification of transcriptionally responsive genes to EN2 and the experimental design added a level of confidence that we would be more likely to identify true targets. Two physiologically relevant breast cancer cell lines were employed so we could discern targets that were cell line specific. This was key as there were numerous differentially expressed genes that were unique to each cell line. Additionally, the levels of ectopic *EN2* expression were comparable to that documented in MDA-MB-435S cells, so we could rule out dosage responsive targets found only at non-physiological elevated levels.

Interestingly, several molecular chaperone family members were induced in response to *EN2* expression and these proteins fittingly play a key role in promoting survival and in maintaining the malignant phenotype in tumor cells. Heat shock proteins and their co-chaperones have recently been shown to regulate both apoptosis and cell death. Hsp70, found to be highly expressed in response to *EN2*, prevents apoptosome formation, diminishes Apaf-1-independent apoptosis and inhibits caspase-dependent events that occur later in apoptosis [12]. Both Hsp90 and Hsp70 bind and stabilize the survival protein Akt [11,13]. The induction of Hsp40/DnaJ, another key regulator of apoptosome formation [12], also coincided with *EN2* epression in our microarray study. Although molecular chaperone proteins are consistently expressed at higher levels in tumor cells, they are not always found to be further induced in transcription profiles of other microarray studies using oncogenes (http://clarkelabs.georgetown.edu/BreastStudies), suggesting that they are functional *EN2* targets and not a mere consequence of the transformed phenotype in our study.

One chaperone in particular, Hsp90, is found at elevated levels in cancer cells where it regulates the function and stability of key proteins that protect cancer cells from the stress of chemotherapy, hypoxia and their own inherent genetic instability [9]. The up-regulation of such protective proteins would be most advantageous for cancer cells to thrive and suggests that *EN2* may promote mammary hyperplasia and tumorigenesis by providing both a proliferative and survival advantage. Mammary epithelial cells that express *EN2* are quite sensitive to 17-AAG and the cessation of proliferation, following exposure to 17-AAG *in vitro*, is accompanied by cell death. This observation implies that *EN2* is truly an oncogene that transforms these cells, in part through the upregulation of Hsp90, rendering them more sensitive to 17-AAG.

In addition, molecular chaperones have recently been shown to be required for GR and PR nuclear mobility and the Hsp90 inhibitor, Geldanamycin, inhibits their rapid nuclear trafficking and function [14]. Hsp90 has been shown to stabilize 'client proteins' such as HER2, AKT, steroid receptors, mutated p53 and $p210^{Bcr\text{-}Abl}$, which are often exploited by cancer cells for growth and/or survival advantages [5]. It would be fitting if EN2 is

another oncogenic Hsp90 client protein and/or Hsp90 enhances its localization and activity, as we have clearly shown that *EN2* provides a proliferative advantage to cells.

The downregulation of *HOXC6* is worthy of note but exactly how this could be an oncogenic mechanism for EN2 is not clear. *HOXC6* is often overexpressed in human leukemia yet *Hoxc6* transcripts were found to decrease during pregnancy when the mammary gland is highly proliferative and *HOXC6* is expressed at low levels in human precancerous tissue and not expressed in breast cancer tissue [15]. Another study reports that *Hoxc6* expression in mammary epithelial cells appears to promote adult mammary gland expansion while *Hoxc6* expression in the mammary stroma appears to promote regression and involution of the gland [16]. If aberrant *EN2* expression represses the latter activity of *HOXC6*, *EN2* may supersede the normal growth program and restraints of the mammary gland and allow for immense expansion.

Unlike a screen involving primary breast tissue, the cancer cell lines that the targets were identified in are immediately accessible for further genetic studies. This provides a faster path from gene identification to functional validation and the precise role of a target and its relative contributions to *EN2*-regulated tumorigenesis can be explored initially in these cells. The exact molecular mechanism(s) by which *EN2* fosters tumorigenicity still need to be elucidated, but this is the very first glimpse into potential mammalian targets that may convey the transcriptional signal from *EN2* in human breast cancer. To our knowledge, not only have there been no reports of *EN2* targets that are relevant to cancer, but there has not been a single *EN2* target identified in any human adult tissue. A recent study using chromatin immunoprecipitation with *Drosophila engrailed* however, pulled out direct targets such as *Wnt2*, *Frizzled 2*, *armadillo* (*beta-catenin*), *EGFR*, *EGFR* ligand, *Gsc* (goosecoid), and *p53* [17]. Similar to other homeodomain-containing proteins, the remarkable conservation of the homeodomain makes it difficult to understand how transcriptional specificity can be attained. A probable explanation for homeoprotein specificity is their association with co-factors. Clearly, homeoproteins have been shown to associate with numerous proteins, including members of the same or divergent homeoprotein family and non-homeodomain proteins. Several En-2 co-factors

have recently been identified and belong to several classes of transcriptional regulators; *Groucho*, *Pbx*, *Meis*, *Foxa2* [18,19,20].

Materials and Methods

Cell lines

The MDA-MB-435S human mammary epithelial cell line was derived from a ductal adenocarcinoma and was grown in RPMI 1640 medium supplemented with 10% fetal calf serum (FCS), 10 ng/ml of epidermal growth factor (EGF) and 10 μg/ml of insulin. The T-47D human mammary epithelial cell line was isolated from a ductal adencarcinoma and was grown in Lebowitz-L15 medium supplemented with 10% FCS. The techniques used to engineer T-47D cells to ectopically express *EN2* and *EGFP*, the techniques employed to suppress EN2 protein levels in MDA-MB-435S cells, and the steps involved in RT-PCR analysis have been described previously [1].

RNA Isolation and Oligonucleotide Array Hybridization

Total RNA was isolated from MDA-MB-435S cells 3 days post transfection with either siEN2 or siCTRL and from T-47D cells 3 days post infection with concentrated and filtered VSV supernatant possessing *EN2*- or *GFP*-containing retroviruses, when the cells were >95% *GFP*-positive, with the RNeasy mini kit (Qiagen). Triplicates of RNA for each experimental sample were provided for hybridization. Total RNA quality was assessed by running ~100 ng of total RNA on the Agilent Bioanalyzer. Target labeling was performed as specified by the Affymetrix protocol EukGe Ws2v4 with 10 μg of total RNA. After IVT, 15 ng of the resulting cRNA was used for the fragmentation reaction and 10 μg of the fragmented reaction was hybridized onto the HG-U133A expression microarray (Affymetrix). The quality of the cRNA amplification was also assessed by running 1 μl of the cRNA reaction on the Agilent Bioanalyzer. Staining, washing and scanning were carried out in accordance with Affymetrix protocol EukGe Ws2v4.

Expression Profiling Data Analysis (written by Dr. Le Blanc)

Background-corrected batch-normalized $\log_2$-valued probeset summaries were obtained from probe raw signals using the RMA subroutine [21]. We used the RMA version implemented in R and freely available from the Bioconductor website at www.bioconductor.org/. Gene expression level variance across samples in a given experiment is a function of the experimental design. Since the present design involves two different biological systems which necessarily would impart a higher level of gene variance whenever all samples would be considered at once for gene variance estimatation, we chose to normalize unlogged gene expression levels in a two-step procedure in order to better account for the higher degree of gene variance inherent to the experimental design and better uncover parallel patterns of up- or downregulation across the biological systems. First, in vectorial notation, we refer to a given gene expression profile across all samples of either 435S or T47D biological system as ${}^{435S}\mathbf{g}$ or ${}^{T47D}\mathbf{g}$, respectively. These expression profiles were centered and normalized such that the scalar product ${}^{B}\mathbf{g}_1 \bullet {}^{B}\mathbf{g}_2$ between any two genes in either biological system B = 435S or T47D directly yields their Pearson correlation in that system. Next, a normalized gene expression profile **g** across all samples of the experimental set-up was defined in terms of the outer product $\mathbf{g}=({}^{435S}\mathbf{g} \otimes {}^{T47D}\mathbf{g})/\sqrt{2}$. When defined this way, the scalar product between any two genes

$$\mathbf{g}_1 \bullet \mathbf{g}_2 = ({}^{435S}\mathbf{g}_1 \bullet {}^{435S}\mathbf{g}_2 + {}^{T47D}\mathbf{g}_1 \bullet {}^{T47D}\mathbf{g}_2)/2$$

simply amounts to the Pearson correlation as averaged over the two independent biological systems. Two genes with simultaneously strong positive or negative correlations across both biological systems will therefore retain a high correlation in the combined system.

Biological samples were split into two classes (*EN2+* or *EN2-*) as defined by the presence or absence of EN2. After gene normalization and filtering, differentially expressed genes were identified by first computing their two-sample Welch t-statistics.

$$ {}^{B}t = \frac{{}^{B}\overline{g}_{EN2+} - {}^{B}\overline{g}_{EN2-}}{\sqrt{\frac{\left({}^{B}s_{EN2+}\right)^2}{n_{EN2+}} + \frac{\left({}^{B}s_{EN2-}\right)^2}{n_{EN2-}}}} $$

where ${}^{B}\overline{g}_{EN2\pm}$ denotes the gene class average in the biological system B and ${}^{B}s_{EN2\pm}$ the corresponding gene class standard deviation. When gene expression profiles are normalized as above and when there is an equal number n of samples in each class of each biological system B, it is relatively straightforward to prove that

$$ {}^{B}t = f^{(n)}\left({}^{B}\overline{G}_{EN2+}\right) $$

where

$$ f^{(n)}\left({}^{B}\overline{G}_{EN2+}\right) = \sqrt{2(n-1)} \cdot \frac{{}^{B}\overline{G}_{EN2+}}{\sqrt{(1-{}^{B}\overline{G}^{2}_{EN2+})}} $$

and ${}^{B}\overline{G}_{EN2+} = \sqrt{2n} \cdot {}^{B}\overline{g}_{EN2+}$. For the combined system, one finds

$$ t = f^{(2n)}\left(\frac{1}{2}\left({}^{435S}\overline{G}_{EN2+} + {}^{T47D}\overline{G}_{EN2+}\right)\right) $$

A gene with high ${}^{B}t$-score, that is, ${}^{B}\overline{G}_{EN2+} \to \pm 1$ across both biological systems simultaneously will therefore retain a high t-score in the combined system. Note that ${}^{B}\overline{G}_{EN2+} \to \pm 1$ if and only if the normalized gene expression profile across all samples approaches the idealized down/up expression pattern

$$ \frac{1}{\sqrt{2n}}\left(\underbrace{\mp 1 \mp 1 \ldots \mp 1}_{n}\underbrace{\pm 1 \pm 1 \ldots \pm 1}_{n}\right), $$

whenever samples are ordered according to the layout (*EN2-* samples, *EN2+* samples). Statistical significance was assessed by computing t-score associated p-values, adjusted for multiple comparisons. These adjusted p-values were estimated by permutation

following Westfall and Young's step-down maxT adjusted *p*-value procedure [22] using the mt.maxT subroutine from the multtest R package available at the Bioconductor repository web site [23]. A subset of analyses, namely observed vs expected frequencies of Gene Ontology (GO) classes, were performed using BRB-Array tools v3.1 developed by Dr. Richard Simon and Amy Peng Lam (http://linus.nci.nih.gov/BRB-ArrayTools.html).

In order to obtain clusters of a manageable size, a filter was applied in order to select genes with pre-determined changes in the presence or absence of the oncogene *EN2*. We elected to filter genes for which the unlogged RMA values increased or decreased by at least 2^5=32 units on average in presence of the oncogene *EN2* when compared to a baseline where *EN2* is either not expressed or its mRNA is interfered with to disrupt protein translation. Note that this filter does not specify a pre-determined multiplicative fold-change threshold. Rather, it reflects the fact that currently favored background noise models for Affymetrix oligonucleotide microarrays assume an additive contribution of noise (N) to signal (S), and that N will mostly affect signals in the low intensity range, that is, signals for which $\log_2 S \approx 5\text{-}7$. 1417 out of the 22215 probesets passed the filter. The list of all 1417 probesets that passed the filter along with their expression values and adjusted *p*-values are available as Supplemental Data in the Sauvageau lab server. Hierarchical clustering of the first 50 ranked differentially expressed filtered genes was performed using Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster) and the resulting clusters were visualized using Mapple Tree [24] (http://sourceforge.net/projects/mapletree).

Acknowledgements

We thank Dr. Alexandra Joyner (NYU, NYC) for the *En-2* cDNA; Dr. Pearl Campbell (OHRI, Ottawa) for performing the Affymetrix hybridization; and Dr. Morag Park (McGill University, Montréal) for the MDA-MB-435S cell line. This work was supported by funds from the NCIC (15064) to G.S. N.M. is a recipient of a CIHR research award, G.S. is a scholar of the Leukemia and Lymphoma Society of USA.

Table 4.1 Subset of genes that were significantly ($P < 0.05$) up-regulated ≥ 2 fold or down-regulated ≤ 0.5 fold in response to *EN2* expression in either MDA-MB-435S or T-47D cell lines with the corresponding fold change observed in the other cell line indicated. Bold font indicates transcripts which were induced at comparable levels in both cell lines.

Fold change in MDA-MB-435S	Gene description	Fold change in T-47D
0.90	*heat shock 70kDa protein 6 (HSP70B')*	33.45
2.02	***prostate differentiation factor***	**7.19**
2.02	***protein kinase H11***	**3.97**
2.02	***tumor necrosis factor, alpha-induced protein 3***	**3.84**
1.10	*DNA-damage-inducible transcript 3*	3.81
1.30	*crystallin, alpha B*	3.79
0.90	*BCL2-associated athanogene 3*	3.65
1.10	*heat shock 105kD*	3.48
3.39	*cyclin B1*	0.84
1.10	*S100 calcium binding protein P*	3.33
3.26	*baculoviral IAP repeat-containing 5 (survivin)*	0.91
0.90	*heat shock 70kDa protein 1A*	3.22
3.20	*basic leucine zipper and W2 domains 2*	1.11
1.20	*glycoprotein (transmembrane) nmb*	2.94
1.10	*ferritin, light polypeptide*	2.92
0.90	*homocysteine-inducible, ubiquitin-like domain member 1*	2.86
1.00	*ESTs, Highly similar to FRIL_HUMAN Ferritin light chain*	2.79
2.66	*protein disulfide isomerase-related protein*	1.44
1.00	*kynurenine 3-monooxygenase (kynurenine 3-hydroxylase)*	2.51
1.20	*solute carrier family 3, member 2*	2.48
0.90	*solute carrier family 7, member 11*	2.48
2.42	*hyaluronan-mediated motility receptor (RHAMM)*	0.83
0.90	*H4 histone family, member H*	2.41
2.40	*protein regulator of cytokinesis 1*	0.96
2.38	***tumor rejection antigen (gp96) 1***	**2.04**
1.00	*Rag D protein*	2.35
2.34	*immediate early response 3*	1.03
1.00	*putative transmembrane protein*	2.29
2.27	*topoisomerase (DNA) II alpha 170kDa*	0.86
2.26	*RAB6 interacting, kinesin-like (rabkinesin6)*	0.84
2.22	*V-ATPase C2 subunit*	1.46
2.21	*staufen, RNA binding protein (Drosophila)*	1.17
2.20	***heat shock 70kDa protein 5 (78kDa)***	**1.99**
2.19	*sorting nexin 4*	1.11
2.15	*hypothetical protein FLJ10540*	0.77
2.14	*Bcl-2-associated transcription factor*	1.39
1.10	*hypothetical protein FLJ20059*	2.13
1.40	*dual specificity phosphatase 1*	2.12
2.11	*epithelial cell transforming sequence 2 oncogene*	0.81
1.40	*DnaJ (Hsp40) homolog, subfmaily B, member 1*	2.10
1.20	*ornithine decarboxylase 1*	2.09
2.08	*eukaryotic translation initiation factor 3, subunit 6 48kDa*	1.01
2.06	*Drosophila discs large-1 tumor supressor-like*	0.79
0.90	*3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2*	2.06
1.00	*heat shock protein (hsp110 family)*	2.05
1.00	*SH3 domain binding glutamic acid-rich protein*	2.05
1.00	*activating transcription factor 3*	2.04
2.01	*palmitoyl-protein thioesterase 1 (ceroid-lipofuscinosis)*	1.33
1.00	*meningioma expressed antigen 5 (hyaluronidase)*	2.01
0.49	*ADP-ribosylation factor 3*	0.74
0.48	*H1 histone family, member 2*	1.60

Fold change in MDA-MB-435S	Gene description	Fold change in T-47D
0.47	*immunoglobulin superfamily, member 3*	1.18
0.47	*ubiquinol-cytochrome c reductase binding protein*	1.12
0.45	*hypothetical protein FLJ11848*	1.31
0.43	*secretory carrier membrane protein 1*	1.01
0.43	*hypothetical protein FLJ11193*	1.26
0.43	*protein tyrosine phosphatase type IVA, member 2*	0.90
0.42	*START domain containing 7*	1.09
0.40	*actin related protein 2/3 complex, subunit 2, 34kDa*	0.87
0.40	*fasciculation and elongation protein zeta 2 (zygin II)*	1.16
0.40	*huntingtin interacting protein 2*	1.08
0.39	*hypothetical gene supported by AF038182; BC009203*	1.62
0.37	*conserved gene amplified in osteosarcoma*	0.94
0.70	*serine (or cysteine) proteinase inhibitor*	0.34
0.33	*methylene tetrahydrofolate dehydrogenase*	1.15
0.32	*APG12 autophagy 12-like (S. cerevisiae)*	1.07
0.27	*PAI-1 mRNA-binding protein*	1.05
0.27	*serum-inducible kinase*	1.22

Table 4.2 Subset of regulated transcripts common to both cell lines with significant induction or repression in response to *EN2* expression with adjusted P values less than 0.05.

Affy ID	Gene	Gene Description (Genes upregulated in response to EN2)	P value
207668_x_at	*P5*	*protein disulfide isomerase-related protein*	0.0022
208639_x_at	*P5*	*protein disulfide isomerase-related protein*	0.0043
202655_at	*ARMET*	*arginine-rich, mutated in early stage tumors*	0.0065
200599_s_at	*TRA1*	*tumor rejection antigen (gp96) 1*	0.0087
221923_s_at	*NPM1*	*nucleophosmin (nucleolar phosphoprotein B23, numatrin)*	0.0087
200691_s_at	*HSPA9B*	*heat shock 70kDa protein 9B (mortalin-2)*	0.0087
212510_at	*KIAA0089*	*KIAA0089 protein*	0.0087
200666_s_at	*DNAJB1*	*DnaJ (Hsp40) homolog, subfmaily B, member 1*	0.0087
221577_x_at	*PLAB*	*prostate differentiation factor*	0.0087
202934_at	*HK2*	*hexokinase 2*	0.0087
216640_s_at	*P5*	*protein disulfide isomerase-related protein*	0.0087
202636_at	*NEDF*	*zinc finger protein 103 homolog (mouse)*	0.0087
216450_x_at	*TRA1*	*tumor rejection antigen (gp96) 1*	0.0087
202130_at	*RIOK3*	*sudD suppressor of bimD6 homolog (A. nidulans)*	0.0108
202382_s_at	*GNPI*	*glucosamine-6-phosphate isomerase*	0.0108
209694_at	*PTS*	*6-pyruvoyltetrahydropterin synthase*	0.0152
218088_s_at	*RRAGC*	*Rag C protein*	0.0173
218574_s_at	*LMCD1*	*LIM and cysteine-rich domains 1*	0.0173
201952_at	*ALCAM*	*activated leukocyte cell adhesion molecule*	0.0173
214328_s_at	*HSPCA*	*heat shock 90kDa protein 1, alpha*	0.0173
200825_s_at	*HYOU1*	*hypoxia upregulated 1*	0.0173
204255_s_at	*VDR*	*vitamin D (1,25- dihydroxyvitamin D3) receptor*	0.0173
209295_at	*TNFR*	*tumor necrosis factor receptor superfamily, member 10b*	0.0238
201925_s_at	*DAF*	*decay accelerating factor for complement*	0.0238
209161_at	*PRPF4*	*PRP4 pre-mRNA processing factor 4 homolog (yeast)*	0.0238
202302_s_at	*FLJ11021*	*hypothetical protein FLJ11021 similar to splicing factor*	0.0238
204266_s_at	*CHK*	*choline kinase*	0.0238
209238_at	*STX3A*	*syntaxin 3A*	0.0238
208852_s_at	*CANX*	*calnexin*	0.0238
200598_s_at	*TRA1*	*tumor rejection antigen (gp96) 1*	0.0260
200068_s_at	*CANX*	*calnexin*	0.0260
218145_at	*C20orf97*	*chromosome 20 open reading frame 97*	0.0260
210275_s_at	*ZNF216*	*zinc finger protein 216*	0.0260
200732_s_at	*PTP4A1*	*protein tyrosine phosphatase type IVA, member 1 / PRL-1*	0.0260
208813_at	*GOT1*	*glutamic-oxaloacetic transaminase 1)*	0.0281
221691_x_at	*NPM1*	*nucleophosmin (nucleolar phosphoprotein B23, numatrin)*	0.0281
218358_at	*MGC11256*	*hypothetical protein MGC11256*	0.0281
212501_at	*CEBPB*	*CCAAT/enhancer binding protein (C/EBP), beta*	0.0281
203241_at	*UVRAG*	*UV radiation resistance associated gene*	0.0325
211969_at	*HSPCA*	*heat shock 90kDa protein 1, alpha*	0.0325
203534_at	*LSM1*	*Lsm1 protein*	0.0368
212724_at	*ARHE*	*ras homolog gene family, member E*	0.0368
200790_at	*ODC1*	*ornithine decarboxylase 1*	0.0368
209180_at	*RABGGTB*	*Rab geranylgeranyltransferase, beta subunit*	0.0368
203159_at	*GLS*	*glutaminase*	0.0368
208581_x_at	*MT1X*	*metallothionein 1X*	0.0390
210849_s_at	*VPS41*	*vacuolar protein sorting 41 (yeast)*	0.0390
211936_at	*HSPA5*	*heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)*	0.0390
209681_at	*SLC19A2*	*solute carrier family 19 (thiamine transporter), member 2*	0.0411
212434_at	*GRPEL1*	*GrpE-like protein cochaperone*	0.0411
202106_at	*GOLGA3*	*golgi autoantigen, golgin subfamily a, 3*	0.0411
221841_s_at	*KLF4*	*Kruppel-like factor 4 (gut)*	0.0411
203511_s_at	*BET3*	*similar to yeast BET3 (S. cerevisiae)*	0.0433

Affy ID	Gene	Gene Description (upregulated in response to EN2 *cont*)	p value
210211_s_at	*HSPCA*	*heat shock 90kDa protein 1, alpha*	0.0455
218233_s_at	*C6orf49*	*over-expressed breast tumor protein*	0.0476
208638_at	*ATP6V1C2*	*V-ATPase C2 subunit*	0.0476
208726_s_at	*EIF2S2*	*eukaryotic translation initiation factor 2*	0.0476
219212_at	*HSP70-4*	*ortholog of mouse heat shock protein, 70 kDa*	0.0476
201453_x_at	*RHEB*	*Ras homolog enriched in brain 2*	0.0498

Affy ID	Gene	Gene Description (Genes downregulated in response to EN2)	P value
208691_at	*TFRC*	*transferrin receptor (p90, CD71)*	0.0476
209366_x_at	*CYB5*	*cytochrome b-5*	0.0281
209154_at	*TIP-1*	*Tax interaction protein 1*	0.0238
200809_x_at	*RPL12*	*ribosomal protein L12*	0.0173
217776_at	*RDH11*	*androgen-regulated short-chain dehydrogenase/reductase 1*	0.0173

Table 4.3 Gene Ontology (GO) Mining Tool provided predictions that 1.01 genes would be associated with heat shock protein activity and 4.25 genes with chaperone activity among the first 477 probesets which were significant at the nominal 0.001 level of the univariate two-sample t-test.

GO id	GO classification	Observed in selected subset	Expected in selected subset	Observed/ Expected
0003773	heat shock protein activity	7	1.01	6.92
0003754	chaperone activity	22	4.25	5.17
0016860	intramolecular oxidoreductase activity	5	1.1	4.53
0015078	hydrogen ion transporter activity	10	3.38	2.96
0015077	monovalent inorganic cation transporter activity	10	3.75	2.67
0003925	small monomeric GTPase activity	11	4.34	2.53
0004197	cysteine-type endopeptidase activity	5	2.07	2.42
0003924	GTPase activity	13	5.52	2.36
0016616	oxidoreductase activity\, acting on the CH-OH group of donors\, NAD or NADP as acceptor	6	2.67	2.25
0050800	hydrolase activity\, acting on acid anhydrides\, acting on GTP\, involved in cellular and subcellular movement	11	4.96	2.22
0008248	pre-mRNA splicing factor activity	6	2.71	2.21
0003714	transcription corepressor activity	7	3.19	2.19
0016614	oxidoreductase activity\, acting on CH-OH group of donors	6	2.76	2.18
0008324	cation transporter activity	13	6.23	2.09
0005525	GTP binding	12	5.84	2.06
0019001	guanyl nucleotide binding	12	5.93	2.02

References

1. Martin,N.L. et al. *EN2* is a candidate oncogene in human breast cancer. *Oncogene.* (2005).

2. Ashburner,M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat.Genet.* **25**, 25-29 (2000).

3. Liu,G. et al. NetAffx: Affymetrix probesets and annotations. *Nucleic Acids Res.* **31**, 82-86 (2003).

4. Kikuchi,M. et al. Functional analysis of human P5, a protein disulfide isomerase homologue. *J.Biochem.(Tokyo).* **132**, 451-455 (2002).

5. Isaacs,J.S., Xu,W. & Neckers,L. Heat shock protein 90 as a molecular target for cancer therapeutics. *Cancer Cell.* **3**, 213-217 (2003).

6. Eustace,B.K. et al. Functional proteomic screens reveal an essential extracellular role for hsp90 alpha in cancer cell invasiveness. *Nat.Cell Biol.* **6**, 507-514 (2004).

7. Luparello,C., Sirchia,R. & Pupello,D. PTHrP [67-86] regulates the expression of stress proteins in breast cancer cells inducing modifications in urokinase-plasminogen activator and MMP-1 expression. *J.Cell Sci.* **116**, 2421-2430 (2003).

8. Picard,D. Hsp90 invades the outside. *Nat.Cell Biol.* **6**, 479-480 (2004).

9. Neckers,L. & Lee,Y.S. Cancer: the rules of attraction. *Nature.* **425**, 357-359 (2003).

10. Kamal,A. et al. A high-affinity conformation of Hsp90 confers tumour selectivity on Hsp90 inhibitors. *Nature.* **425**, 407-410 (2003).

11. Basso,A.D. et al. Akt forms an intracellular complex with heat shock protein 90 (Hsp90) and Cdc37 and is destabilized by inhibitors of Hsp90 function. *J.Biol.Chem.* **277**, 39858-39866 (2002).

12. Takayama,S., Reed,J.C. & Homma,S. Heat-shock proteins as regulators of apoptosis. *Oncogene.* **22**, 9041-9047 (2003).

13. Gao,T. & Newton,A.C. The turn motif is a phosphorylation switch that regulates the binding of Hsp70 to protein kinase C. *J.Biol.Chem.* **277**, 31585-31592 (2002).

14. Elbi,C. et al. Molecular chaperones function as steroid receptor nuclear mobility factors. *Proc.Natl.Acad.Sci.U.S.A.* **101**, 2876-2881 (2004).

15. Friedmann,Y., Daniel,C.A., Strickland,P. & Daniel,C.W. Hox genes in normal and neoplastic mouse mammary gland. *Cancer Res.* **54**, 5981-5985 (1994).

16. Garcia-Gasca,A. & Spyropoulos,D.D. Differential mammary morphogenesis along the anteroposterior axis in Hoxc6 gene targeted mice. *Dev.Dyn.* **219**, 261-276 (2000).

17. Solano,P.J. et al. Genome-wide identification of in vivo Drosophila Engrailed-binding DNA fragments and related target genes. *Development.* **130**, 1243-1254 (2003).

18. Alexandre,C. & Vincent,J.P. Requirements for transcriptional repression and activation by Engrailed in Drosophila embryos. *Development.* **130**, 729-739 (2003).

19. Foucher,I. et al. Joint regulation of the MAP1B promoter by HNF3beta/Foxa2 and Engrailed is the result of a highly conserved mechanism for direct interaction of homeoproteins and Fox transcription factors. *Development.* **130**, 1867-1876 (2003).

20. Kobayashi,M. et al. Engrailed cooperates with extradenticle and homothorax to repress target genes in Drosophila. *Development.* **130**, 741-751 (2003).

21. Irizarry,R.A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* **4**, 249-264 (2003).

22. Westfall,P.H., Krishen,A. & Young,S.S. Using prior information to allocate significance levels for multiple endpoints. *Stat.Med.* **17**, 2107-2119 (1998).

23. Dudoit,S., Shaffer,J.P. & Boldrick,J.C. Multiple hypothesis testing in microarray experiments. *Statistical Science.* **18**, 71-103 (2003).

24. Eisen,M.B., Spellman,P.T., Brown,P.O. & Botstein,D. Cluster analysis and display of genome-wide expression patterns. *Proc.Natl.Acad.Sci.U.S.A.* **95**, 14863-14868 (1998).

CHAPITRE 5

Conclusions, Perspectives and Future Directions

Chapter 5 is a discussion of the relevance of the findings presented in this thesis and adresses potential future studies that could be performed to further understand the roles of *En-1* and *EN2* in the regulation of normal mammary gland development and breast cancer, respectively.

En-1 in mouse mammary gland development

Our findings provide some initial insight as to which developmental stage coincides with *En-1* expression and where *En-1* expressing cells appear to reside with respect to the actively growing ductal tree. We have shown that there are some clear candidate cell populations which might express *En-1*, while ruling out others such as the inner most body cells of the TEB and the luminal epithelial cells of the ducts. Our immunohistochemical studies document En-1 expression in putative cap cells, myoepithelial cells, periductal fibroblasts and in a few stromal cells. As one can appreciate from Fig. 5.1 however, we cannot conclude with confidence, from our intial immunohistochemistry data, whether En-1 is expressed in the myoepithelial cells, the thickened basal lamina, or both. In addition, it is possible that the En-1 expression that we perceive as arising solely from the cap cell layer of the TEB, may also arise from some of the body cells in the TEB and/or the thin basal lamina surrounding the TEB (Fig. 5.1).

Given the small window of *En-1* expression and the putative cell types it is exclusively found in, one possibility is that *En-1* expressing cap cells are those progenitors that give rise to myoepithelial cells in the newly forming ducts. Since primary progenitor cells represent one daughter cell of a stem cell division and divide repeatedly before becoming committed to a given lineage, one possibility is that *En-1* may be expressed in the secondary progenitors arising from the primary progenitors that have made the decision to continue along the myoepithelial lineage in the rapidly growing pubertal ductal system. There are no reports of proteins upstream of *En-1*, or known targets of En-1 signaling, that are involved in the specification of the myoepithelial lineage from cap cells within the growing ductal system, but similarly, there are few upstream regulators and even fewer direct targets of En-1 signaling that have been identified in mammalian cells. Interestingly, studies have suggested that Wnt-signaling is involved in myogenic specification [1]. In addition, chromatin immunoprecipitation (ChIP) studies have shown that potential target genes of *Drosophila engrailed* are involved in muscle development pathways [2]. To investigate this possibility, the specific cell type(s) that express *En-1* must first be clearly defined.

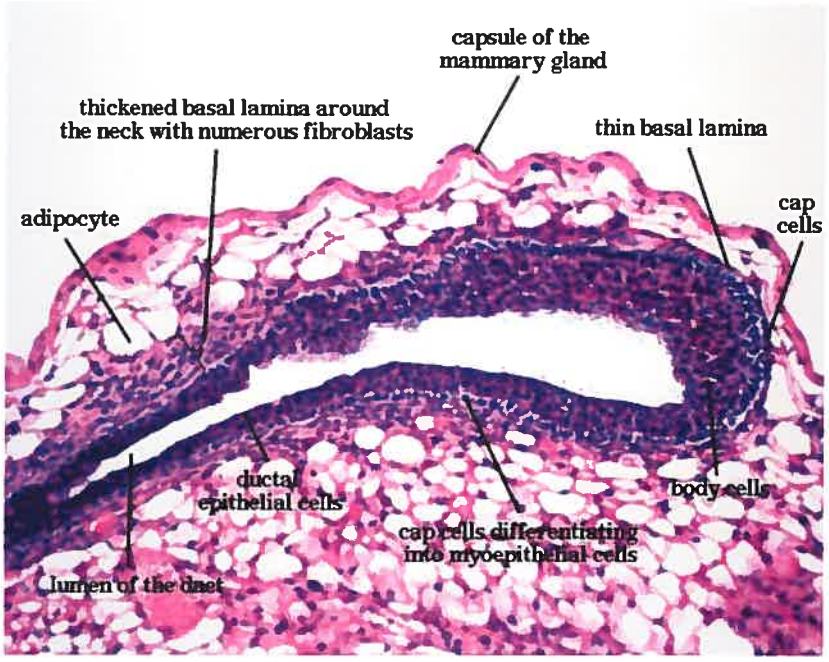

Fig. 5.1 Longitudinal section of a TEB from a 5-week old virgin Balb/c mouse. The TEB within the pubertal mouse mammary gland contains two cell types: body & cap cells. The cap cells at the tip of the growing TEB are highly proliferative and differentiate into the myoepithelial cells, which surround the mature duct. The body cells are the most prominent cells in the TEB and differentiate into the ductal epithelial cells. Tissue section image reprinted from http://mammary.nih.gov/atlas/structures.

To achieve a clearer picture, it will be necessary to perform precise co-localization experiments at different developmental stages and also in available models with impaired ductal development during puberty. Immunohistochemistry employing several markers, in addition to β-gal, all with unique detection agents, will allow for co-localization of En-1 protein with definitive marker expression. Attempts have been made to correlate cell type specific cytokeratins and cell surface markers with *in vitro* differentiative capacity, to identify particular cell lineages and to further determine which cells have stem cell-like properties based on patterns of marker expression. Isolated primary mammary cells expressing *En-1* can be divided into either luminal or myoepithelial/basal populations and the *in vitro* differentiative capacity of these cells in tissue culture can then be tested in an attempt to define multipotential cells such as cap

cells. Epithelial-specific antigen (ESA), mucin 1 (MUC1) and cytokeratin 8 (K8) are surface markers of luminal epithelial cells while CD10, α6-integrin and K14 are surface markers that are specific to myoepithelial cells. Keratin 6 (K6) is normally expressed in the body cells of TEBs during ductal morphogenesis but not in MECs from mammary ducts [3]. These and other markers can be utilized to characterize the cell fraction that expresses *En-1* to determine if they are indeed myoepithelial, stromal, fibroblast, adipocyte and/or cap cell in origin.

Transplant experiments were performed to determine if the mammary gland phenotype documented in *En-1*$^{-/-}$ females is intrinsic to mammary epithelial cells or due to systemic defects in the *En-1*$^{-/-}$ mutant mice. The resulting transplants were able to grow to varying extents, some of them were able to form TEBs and/or were able to, albeit more sparsely, fill the mammary fat pad. Several possibilities may account for these observations, including a systemic hormonal deficiency in the *En-1*$^{-/-}$ mutant mice, the loss of *En-1* affecting a particular organ which influences the mammary gland, and/or the mutants suffer from a mere delayed puberty.

The fact that *En-1* expression coincides with the onset of puberty also raises the possibility that *En-1* may be produced in response to estrogen to carry out the hormonal stimulation of the mammary epithelium. No defects were apparent in whole mounts derived from *En-1*$^{-/-}$ mutant mice before 3 weeks of age, demonstrating that En-1 is dispensible for the development of the ductal tree preceeding puberty. However, when analyzed at 3 ½ weeks of age, mutant glands displayed delayed ductal outgrowth and TEB formation. Another possibility is that *En-1* may regulate hormone production. Initial expression of *En-1* commences at day 15, which slightly precedes the onset of puberty, and the retarded growth seen in the *En-1* null mutant mouse mammary gland may be a result of the lack, or reduced production, of hormones that normally initiate this growth phase. This will also need to be addressed. The link between hormonal signals and *En-1* expression may come from immunohistochemistry experiments that colocalize En-1 and estrogen R expression in mammary epithelial cells. If this is the case, it will be interesting to see if *En-1* expression can be induced by estrogen in

ovariectomized animals and whether *En-1* is reduced in $ER^{-/-}$ epithelium. Although this might establish whether estrogen activates *En-1* expression, it would remain to be established whether this occurs as a direct or indirect mechanism. In this respect, it will be important to determine whether the *En-1* promoter contains binding sites for estrogen that mediate estrogen-induced transcriptional activation.

Interestingly, studies in normal mammary tissue have shown that progesterone induces *Wnt-4* expression in the mouse mammary gland and that Wnt signaling is the likely mediator of the progesterone signal during pregnancy, a very proliferative stage in postnatal mammary gland development [4]. It is presently thought that Wnt signals also have an important role in the maintenance of stem cell compartments [5]. The reported expression of other Wnt members in unique temporal and spatial patterns during the normal development of the mouse mammary gland provides several candidate upstream regulators of *En-1* expression in non-cancerous breast tissue [6]. Although we showed that *En-1* expression was not consistently activated by etopic *Wnt-1* expression in the mouse mammary gland, future studies will need to address whether *En-1* is downstream to any of the remaining Wnt family members that are expressed in the normal pubertal mammary gland.

It is also possible that compensatory ectopic expression of *En-2* in the C57Bl/6J mammary gland may partially rescue the mammary gland phenotype in $En\text{-}1^{Lki/Lki}$ mutant mice, analogous to the finding that *En-2* expression is likely responsible for the less severe cerebellar brain phenotype when the *En-1 Lki* gene is introduced into the C57Bl/6J strain [7]. Although we have looked at *En-2* expression in mammary tissue from both CD-1 and C57Bl/6J strains, *En-2* expression was never tested in the actual rare $En\text{-}1^{Lki/Lki}$ homozygous mutants that arose. This is something that should be tested to either further suggest or rule out this possibility.

Our studies show that *En-1* is expressed in the mammary epithelium. Although we observed *En-1* expression in what appeared to be periductal fibroblasts and some outlying solitary stromal cells, these results are less convincing and preliminary and will

need to be extended. In addition, we have not verified whether stromal cells expressing *En-1* are only found in the vicinity of the growing ductal tree or whether they are also found in the distal portion of the fat pad where the ductal tree has not yet penetrated. Although this epithelial-free portion of the early pubertal female fat pad did not express detectable levels of En-1 protein by western blot analysis, this technique may not be sensitive enough to detect a small number of solitary stromal cells expressing *En-1*. If we consider that the preliminary western analysis results are indeed reflective of a few *En-1* expressing stromal cells residing adjacent to the mammary epithelium only, this would support the scenario where nearby *En-1* expression in the mammary epithelium could be involved in the crosstalk between the stroma and the epithelial compartments.

Homeodomains and homeoproteins can be internalized by live cells, and a small proportion of Engrailed has been found to associate with membrane regions implicated in signal transduction and secretion and suggest that Engrailed might act as a secreted polypeptidic messenger [8]. These studies suggest the possibility that En-1, a TF, can be converted to a secreted signal, that could control the fate of adjacent cells in a paracrine fashion. Perhaps a portion of En-1 protein produced in the mammary epithelium is secreted and affects the nearby stromal cells and fosters a microenvironment that regulates the ductal expansion during puberty.

Paracrine interactions between individual mammary epithelial cells and/or the adjacent stroma can by investigated by the introduction of a tag that can be visualized in one of the epithelial cell components in a mammary epithelial transplant so it is possible to distinguish the origin of individual cells in mixed epithelial transplants, where transplants contain a mixture of epithelial cells from different origins. If *En-1* expressing cells express a tag, these mammary epithelial cells can be mixed with *En-1* null epithelial cells. If *En-1* is not required in the stromal compartment and the stroma does not contribute to the regeneration of mammary epithelial structures, this is indicative of a pararine mechanism, confined to mammary epithelial cells, of En-1 action. The reciprocal experiment, where *En-1* null cells express a tag, could also be generated.

Although the evidence suggests that En-1 may be involved in ductal growth and morphogenesis during pregnancy, little is known about direct or indirect downstream mediators. Not only should future studies identify and investigate some of these En-1 targets that are specific to normal mammary gland development, but it will be interesting to identify what targets are common to ectopic *EN2* expression during breast tumorigenesis.

EN2 in breast tumorigenesis

Epigenetic activation, where an alteration in gene activity influences the breast tissue phenotype without affecting the genotype, is one plausible explanation for the ectopic expression of *EN2*. Changes in promoter methylation may be involved. Another possibility is that a suppressor of *EN2* (ex. Groucho), which normally serves to maintain repression of *EN2* in certain tissues, may be inactivated and cause the misexpression of *EN2* in breast tissue. Another possible mechanism of ectopic *EN2* expression would be the result of an increase in β-catenin. β-catenin forms a complex with the transcription factor TCF, converting TCF from a transcriptional repressor to an activator of downstream targets and could thereby stimulate the expression of *En-2* without any changes being evident at the genetic level such as rearrangement. TCF has already been shown to activate *c-Myc* in breast tumorigenesis and *EN2* may be activated in addition. Like other homeobox genes, the ectopic expression of wildtype *EN2*, rather than a mutated or altered form, may be involved in the onset of breast carcinogenesis.

Although many studies have reported the deregulated expression of homeobox genes in cancer, relatively few have established direct functional roles for particular homeobox genes in oncogenesis. It is conceivable that in some of these reported cases, the deregulated expression of homeobox genes might be consequential, rather than functionally relevant, for carcinogenesis. The ectopic expression of *EN2* was not merely found to correlate with the progression of breast cancer and with cell lines derived from breast tumors. Its transforming properties *in vitro*, the ability to induce adenocarcinomas *in vivo* and the growth suppressive effect when its expression is knocked down by RNA interference in a breast tumor derived cell line all support a bona fide role for EN2 in

promoting transformation and carcinogenesis. Whether the ectopic expression of *EN2* is an initial and/or causal event in the progression of breast cancer, a downstream and low penetrance event resulting from the initiation of carcinogenesis and/or a result of the genetic instability that occurs, or whether it is pivotal in a later stage such as metastasis, remains to be elucidated. Similar to other homeobox genes, although *EN2* has clear oncogenic properties, its precise function in carcinogenesis in specific tumor types has not been identified. *EN2* may be more fittingly described as a tumour modulator rather than an oncogene at this stage.

A comprehensive analysis of the expression pattern of *EN2* in several types of breast cancer classifications, as well as in other solid tumors such as brain cancer is warranted. Ectopic expression of *EN2* in brain carcinoma would fit the notion of other homeobox genes that are re-expressed in the tissue where they play a pivotal role during embryonic development. EST data has indeed shown that *EN2* expression has been documented in medulloblastoma [9].

Our initial studies looked at 23 primary human breast tumor samples. Of the two *EN2*-positive tumors, one was an adenocarcinoma while the second was an infiltrating ductal carcinoma, inflammatory type. Of the seven *EN2*-positive human cell lines, four were derived from adenocarcinomas, one was derived from a ductal carcinoma, and two were designated as hyperplastic fibrocystic breast tissue. Further studies documented *EN2* expression in 4/59 additional tumor samples, all four *EN2*-positive tumors being classified as infiltrating ductal carcinomas. Thus, *EN2* was shown to be ectopically expressed in a subset of human breast cancer representing approximately 7% in the population tested. A more precise frequency of ectopic *EN2* expression in a much larger sample pool of breast tumors and any possible correlation with a specific subset(s) of tumor phenotypes will need to be screened. Employing a technique of larger scale that is much faster than the RT-PCR/Southern hybridization approach used in these studies will allow one to screen far more samples in a shorter time. Any positive samples can be further analyzed by immunohistochemistry to confirm that the relevant cells within the lesion show nuclear *EN2* staining. To this end, the specificity of the antibody that

recognizes both mouse En-1/human EN1 and mouse En-2/human EN2 has been shown extensively in the literature.

Infiltrating ductal carcinoma (IDC), also known as invasive ductal carcinoma, is the most common type of breast cancer, accounting for 80% of breast cancer diagnoses. IDC begins in the milk ducts of the breast and penetrates the wall of the duct, invading the fatty tissue of the breast and eventually other regions of the body. Inflammatory breast cancer, on the other hand, is quite rare, accounting for only 1% of breast cancer cases. It is associated with the appearance of inflamed breasts with dimples and/or thick ridges caused by cancer cells blocking lymph vessels or channels in the skin over the breast, is extremely aggressive and is associated with poor prognosis. It will be important to explore whether EN2 expression is found in more of these latter samples, potentially providing an early detection marker of this aggressive form of breast cancer, or whether its ectopic expression represents an occurrence that is common to most forms of the disease.

Researchers involved with the Breast Cancer Functional Genomics Group at McGill have been using Laser Capture Microdissection (LCM) to dissect tumor cells from amidst the heterogeneous tumor section. They have agressively been building this tissue bank and the samples contained within would be invaluable to truly ascertain which cells express *EN2* after its expression is linked to a specific subset of breast cancer.

Isolating MSC and progenitor cell-specific promoters and using them to direct expression of *EN2* will enable answering the question of whether EN2 activity is required in stem cells or progenitor cells for *EN2*-positive breast tumor formation. It will be interesting to see if *EN2*-induced tumors contain cells that have myoepithelial, as well as luminal, differentiation, indicating that they might originate in a pluripotent stem or transit cell. Disrupting *EN2* and EN2 target genes of interest within tumor models using RNA interference will allow the potential effects of anti-EN2 therapy to be tested in various mammary tumor models.

Perspectives

There are many examples in which the aberrant expression of homeobox genes that normally regulate growth and development have been implicated in carcinogenesis. For some homeoproteins, it is thought that the wildtype, rather than a mutant form of the protein, produces these oncogenic activities. This implies that the oncogenic activity of some homeoproteins is not due to new or alternative properties, but is a result of their normal functions being carried out in the wrong cellular context. The finding that *En-1* is expressed during the expansion that takes place in pubertal development in the normal mouser mammary gland, combined with the observation that wildtype *EN2* is ectopically expressed in human breast cancer fits into this latter category. Interestingly, both *En-1* and *En-2* behave like oncogenes when expressed in mammary epithelial cells and both inhibit differentiation, yet expression of *En-2* consistently elicited higher measurements in *in vitro* transformation assays compared to *En-1*. Non-transformed mammary epithelial cells that have been engineered to ectopically express wildtype *En-2*, form mammary adenocarcinomas when injected into the cleared fat pads of syngeneic hosts, further supporting the above notion, where the wildtype protein expressed in the wrong cellular context can have oncogenic activity. A key distinguishing feature is that while only *En-1* is expressed in the normal mammary gland, its paralog *EN2* is found to be ectopically expressed in breast carcinogenesis, rather than the re-expression of *En-1*. However, the two paralogs share many conserved functions during development, with *En-2* being able to rescue all but one of the *En-1* null phenotypes, and *En-2* actually appears to exhibit more oncogenic activity in culture and in tumor growth compared to *En-1*.

Importantly, the above findings raise the intriguing possibility that the oncogenic properties of *EN2* in breast tissue might reflect an erroneous extension of the normal postnatal functions of *En-1* in the mammary gland, where EN2 aberrantly mimics the proliferative effects of En-1, the natural inductive signal in mammary epithelium and perhaps leads to the maintenance of an undifferentiated state in breast cells. These observations prompt questions about the basis of En signaling specificity. It will be

instructive to determine which downstream targets are affected by En-1 and EN2 as the two paralogs may have distinct targets in addition to several common targets. Alternatively, the two paralogs could activate different signaling pathways or the same signaling pathways to different extents through differential interactions with co-factors, thereby eliciting unique cellular responses. A genetic understanding of En-1 and EN2 signaling and how they regulate mammary epithelial growth and morphogenesis will improve our ability to manipulate these processes and thus allow us to define strategies for the prevention and treatment of breast cancer.

Conclusion

In summary, the research described in this thesis documents the expression of *Engrailed (En)* genes in the mouse mammary gland and breast cancer tissue and provides evidence that suggests they play a role in both regulating mammary gland development and mammary epithelial cell proliferation in breast carcinogenesis. Hence, these studies reveal that *En* genes have not only acquired a role in the postnatal development of the mouse mammary gland, but in addition, they have evolved to contribute to breast tumorigenesis. These above findings have broad implications in the fields of developmental biology and cancer.

References

1. Seale,P., Polesskaya,A. & Rudnicki,M.A. Adult stem cell specification by Wnt signaling in muscle regeneration. *Cell Cycle.* **2**, 418-419 (2003).

2. Solano,P.J. et al. Genome-wide identification of in vivo Drosophila Engrailed-binding DNA fragments and related target genes. *Development.* **130**, 1243-1254 (2003).

3. Grimm,S.L. & Rosen,J.M. The role of C/EBPbeta in mammary gland development and breast cancer. *J.Mammary.Gland.Biol.Neoplasia.* **8**, 191-204 (2003).

4. Brisken,C. et al. Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. *Genes Dev.* **14**, 650-654 (2000).

5. Hennighausen,L. & Robinson,G.W. Information networks in the mammary gland. *Nat.Rev.Mol.Cell Biol.* **6**, 715-725 (2005).

6. Wong,G.T., Gavin,B.J. & McMahon,A.P. Differential transformation of mammary epithelial cells by Wnt genes. *Mol.Cell Biol.* **14**, 6278-6286 (1994).

7. Bilovocky,N.A. et al. Factors in the genetic background suppress the engrailed-1 cerebellar phenotype. *J.Neurosci.* **23**, 5105-5112 (2003).

8. Joliot,A. et al. Identification of a signal sequence necessary for the unconventional secretion of Engrailed homeoprotein. *Curr.Biol.* **8**, 856-863 (1998).

9. Martin,N.L. et al. *EN2* is a candidate oncogene in human breast cancer. *Oncogene.* (2005).