

# Acknowledgment

This work presented in this thesis was carried out at the Department of Genetics, Institute for Cancer Research at the Norwegian Radium Hospital from August 2013 to May 2014 for the Master's degree in Biotechnology at Norwegian University of Life Science (NMBU).

I am thankful to Professor Anne-Lise Børresen-Dale, the head of the Department of Genetics, for opening the research facilities for master students.

I would like to thank the Research Scientist Therese Sørli for welcoming me to her research group, being an excellent supervisor, always encouraging and for great discussions. I am grateful for having been a part of your research group.

This thesis would not have been possible without postdoctoral Jens Henrik Norum. I am deeply grateful for have been including in your project. I thank you for your invaluable suggestions, advice, patience, discussion and always taking time to answer questions. Your professional dedication and enthusiasm is inspiring.

I would like to thank postdoctoral Charlotte Ramstad Kleiveland at NMBU for being my supervisor and help with the writing of this thesis.

I will express my gratitude to Veronica Skarpeteig, Phoung Vu and Eldri Undlien Due for their expertise and helping me in the laboratory and answering countless questions with pleasure. I would also like to thank the research associate Daniel Nebdal for help with R package and general data problems. I would also like to thank everybody at the Department of Genetics for your positive attitude, encouragement and advise.

Thanks to the collaborators Rune Tuftegård and Erik Fredlund at the Department of Biosciences and Nutrition, and Department of Microbiology at Karolinska Institutet, Stockholm, Sweden.

I would like to thank SFI-CAST and The Norwegian Research Council for founding this project.

Finally, I would like to thank my friends and family, especially my mom, for great encouragement and support through this fun, exciting and demanding period.

Oslo, May 15<sup>th</sup> 2014

---

Helene Zell Thime

## Abstract

Breast cancer accounts for one of the most common cancer-related deaths among women. Glioma associated oncoprotein (GLI) 1 is the major effector of the Hedgehog signaling pathway and is involved in normal breast development. In this thesis, studies to identify tumor characteristics and possible additional mutations promoting GLI1 induced mammary gland tumor formation been conducted. To study the effect of increased GLI1 expression on human breast cancer, a mouse model conditionally overexpressing human GLI1 was established. Four series of GLI1 overexpressing mouse models were used in this thesis.

In this thesis GLI1 induced mammary gland tumors were found to be associated with several different tumor characteristics by immunohistochemistry. GLI1 tumors give rise to different tumor types, both luminal and basal-like subtypes. The GLI1 induced tumors were characterized as triple-negative and were also negative for epidermal growth factor receptor (Egfr).

In two of the GLI1 induced tumors, a positive correlation was found between activated Erk, a downstream member of the MAPK signaling pathway and Kras mutation. Ten of 20 mutated genes identified by whole exome sequencing were confirmed by Sanger sequencing of cDNA. Only the Kras mutation was identified as a mutation that could be involved in tumor development. Further analysis and sequencing have to be conducted to validate more mutations that could be involved in tumor development in tumors overexpressing GLI1.

From the primary tumor that arose in the mice, a small piece was inserted in immunodeficient mice and let the tumor grow before another serial transplantation was conducted. This was conducted for more than ten serial transplantations (generations) to study the stability of GLI1 induced mammary gland tumors. The GLI1 tumors stayed for the major part stable throughout the serial transplantations, except for proliferation marker that increased throughout the serial transplantations.

## Sammendrag

Brystkreft er en av de vanligste kreft-relaterte dødsfall hos kvinner. Glioma associated oncoprotein (GLI) 1 er hoved effektoren i Hedgehog signalveien og er involvert i normal brystutvikling. Unormale mengder av GLI1 proteinet har blitt rapportert i noen brystkreft tilfeller. Undersøkelser av ulike brystkreftrelaterte proteiner og mutasjonsanalyser i GLI1 induerte brystsvulster har blitt utført i denne masteroppgaven. For å kunne studere effekten av overuttrykt GLI1 i brystkreft, er det benyttet en musemodell som betinget uttrykker GLI1 genet og utvikler brystsvulster. Fire serier av slike GLI1 musesvulster ble brukt i denne oppgaven for å studere molekylære egenskaper til svulstene under utvikling .

Ved å bruke immunohistokjemi fant vi at de GLI1 induerte svulstene hadde flere ulike egenskaper og er assosiert med både såkalte lumbale og basal-lignende subtyper. Alle svulstene som ble undersøkt ble karakterisert som trippel-negative (negative for ER, PR Ergb2) og de var også negative for epidermal vekstfaktor reseptor (Egfr).

Videre fant vi en positiv korrelasjon mellom svulster som hadde aktiv Erk, et protein i MAPK signalveien, og svulster som viste Kras mutasjon. Validering av 20 utvalgte gener fra eksom-sekvensering med Sanger sekvensering på cDNA identifiserte ti muterte gener som var uttrykt i svulstene. Imidlertid var det bare mutert Kras som ble identifisert som en mutasjon som potensielt var involvert i kreftutviklingen. Videre analyser av svulstene og ytterligere sekvensering må utføres for å kunne validere flere mutasjoner som kan være viktige for GLI1 induert kreftutvikling.

Modellene er et resultat av serietransplantasjon av en opprinnelig GLI1 induerte primær svulster hos transgene mus, men hvor biter av svulstene ble implantert i serier av immunsupprimerte mus. Dette ble gjort over ti generasjoner for å studere stabiliteten av de GLI1 induerte svulstene. GLI1 svulstene holdt seg for det meste stabile gjennom serietransplantasjonene, bortsett fra for en proliferasjons markør som økte i uttrykk utover generasjonene.

## Abbreviations

BRCA1	Breast cancer 1, early onset
BRCA2	Breast cancer 2, early onset
BSA	Bovine Serum Albumin
C57BL/6	C57Black/6
cDNA	Complementary DNA
DAB	3,3'-Diaminobenzidine
DCIS	ductal carcinomas in situ
ddNTPs	Dideoxynucleotide triphosphates
DDT	Dithiothreitol
DHH	Desert hedgehog
DISP	Dispatched
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotides
ds DNA	Double stranded DNA
EGFR	Epidermal growth factor receptor
ER	Estrogen receptor
ER-alpha	Estrogen receptor alpha
ERBB	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog
Fab	Fragment antigen binding
Fc	Fragment crystallizable
GLI	Glioma associated oncoprotein
H&E	Hematoxylin and Eosin
HER2	Human epidermal growth factor 2
Hh	Hedgehog
Ig	Immunoglobulin
IHC	Immunohistochemistry
IHH	Indian hedgehog
K18	Cytokeratin 18
K5	Cytokeratin 5
K6	Cytokeratin 6
LCIS	lobular carcinomas in situ
M.O.M	Mouse-on-Mouse

MAPK (ERK)	Mitogen-activated protein kinase
MAPKK (MEK)	Mitogen-activated protein kinase kinase
MAPKKK (RAF)	MAP kinase kinase kinase
MMTV	Mouse mammary tumor virus
MMTVrtTA	Mouse mammary tumor virus tetracyclin responsive promoter
mRNA	Messenger ribonucleic acid
NOD/SCID	Non-obese diabetic/severe combined immunodeficient
PCR	Polymerase chain reaction
p-ERK1/2	Phosphorylated extracellular signal-regulated protein kinases 1 and 2
PR	Progesterone receptor
PTCH	Patched receptor
PTEN	Phosphatase and tensin homolog
RasGEF ( Sos)	Ras guanine nucleotide exchange factor
RT	Reverse transcriptase
RTK	Receptor tyrosine kinase
rtTA	Tetracycline-Controlled transcriptional activation
SHH	Sonic hedgehod
SMO	Smoothened receptor
SOLiD	Sequencing by oligo ligation detection
TP53	Tumor protein p53
TRE	Tetracyclin responsive promoter
TREGLI1	Tetracyclin responsive promoter Glioma associated oncoprotein 1

# Table of contents

Acknowledgment .....	I
Abstract .....	II
Sammendrag .....	III
Abbreviations .....	IV
Aim.....	5
1.0 Introduction .....	6
1.1 Biological background.....	6
1.1.1 Mammary gland anatomy and development .....	6
1.1.2 Breast cancer incidence and risk factors .....	7
1.1.3 Development and progression of breast cancer .....	7
1.1.4 Signaling pathways .....	9
1.1.5 The hedgehog signaling pathway.....	9
1.1.6 Epidermal growth factor receptor .....	11
1.1.7 Human epidermal growth factor receptor 2 .....	12
1.1.8 MAP-Kinase signaling pathway .....	12
1.1.9 Estrogen receptor and progesterone receptor .....	14
1.1.10 Cytokeratins .....	15
1.1.11 Ki-67 .....	16
1.1.12 Breast cancer subtypes .....	16
1.1.13 Functional properties of mutated genes .....	18
1.1.14 Model systems .....	18
1.1.15 Ethics.....	19
1.2 Methodological background .....	19
1.2.1 Mice .....	19
1.2.2 Hematoxylin and eosin staining.....	20
1.2.3 Immunohistochemistry .....	20
1.2.3.1 Antibody-antigen binding .....	21
1.2.3.2 Avidin-Biotin complex, immunohistochemistry, and indirect method.....	22
1.2.4 Chain termination sequencing (Sanger sequencing) .....	23
1.2.5 Massively parallel sequencing (SOLiD sequencing).....	25
1.2.6 cDNA synthesis .....	26
1.2.7 NanoDrop 1000 spectrophotometer .....	27
1.2.8 Qubit 2.0 fluorometer.....	28
2.0 Materials.....	29
2.1 Mice .....	29
2.2 Reagents and chemicals.....	30

2.3 Commercial kits.....	30
2.4 Solutions prepared in the lab .....	32
2.5 Antibodies.....	35
2.6 Sequencing primers .....	35
2.7 Various equipment.....	35
2.8 Database and software .....	36
3.0 Methods.....	38
3.1 Mouse model .....	38
3.2 Exome sequencing.....	38
3.3 Immunohistochemistry .....	39
3.3.1 Antibody selection .....	39
3.3.2 Fixation, paraffin embedding and sectioning.....	39
3.3.3 Immunohistochemistry protocol .....	39
3.4 DNA isolation.....	42
3.5 RNA isolation .....	42
3.6 cDNA synthesis .....	43
3.7 Sanger sequencing .....	43
3.7.1 Selected genes and primer design .....	43
3.7.2 PCR amplification.....	46
3.7.3 Chain termination sequencing.....	47
3.7.4 Capillary electrophoresis and results processing .....	47
3.8 Statistical analysis/bioinformatics .....	48
4.0 Results .....	50
4.1 Immunohistochemistry .....	50
4.2 Exome sequencing.....	56
4.3 Sanger sequencing .....	57
5.0 Discussion .....	61
5.1 Methodological discussion .....	61
5.1.1 Immunohistochemistry.....	61
5.1.2 DNA and RNA isolation .....	62
5.1.3 cDNA synthesis .....	62
5.1.4 Advantages and disadvantages with Sanger and SOLiD sequencing technologies.	63
5.1.5 Exome sequencing .....	63
5.1.6 Sanger sequencing .....	63
5.1.7 Comparing exome and Sanger sequencing results.....	66
5.2 Biological discussion.....	67
5.2.1 Tumor characterization .....	67

5.2.2 Gene mutations and impact on tumorigenesis .....	68
5.2.3 Tumor stability .....	70
5.3 Further perspectives.....	71
6.0 Conclusion.....	73
Literature .....	74
Appendix A .....	86
Appendix B .....	88



## Aim

This thesis is a part of a larger project by Jens Henrik Norum and Therese Sørli at the Department of Genetics, Institute for Cancer Research, Oslo University Hospital. My thesis was conducted to further investigate and characterize GLI1 induced mammary tumors at the DNA, RNA and protein levels for their application as models for human breast cancer.

**Aim 1:** Characterize GLI1 induced mammary gland tumors according to immunohistochemical markers

**Aim 2:** Study the stability of GLI1 induced mammary tumors over time

**Aim 3:** Validate whole exome sequencing results by sequencing cDNA with Sanger sequencing technology

**Aim 4:** Identify possible somatic mutations that have contributed to mammary gland tumor development

# 1.0 Introduction

The introduction chapter is divided in biological and methodological background.

## 1.1 Biological background

This chapter gives the biological background to the work in this thesis

### 1.1.1 Mammary gland anatomy and development

The primary function of the mammary gland is to provide nutrition to offspring. The mammary gland is composed of relative dynamic tissue and undergoes cycles of growth, development, differentiation and regression during the different stages of mammary gland development. Mammary gland development is divided in three stages; embryonic, pubertal and adult. Humans are born with underdeveloped mammary glands and between birth and puberty the mammary gland development is arrested. In puberty, the ducts elongate and branches and develop into the full adult mammary gland. In adults, mammary gland development is silenced, except for during pregnancy when the alveolus proliferates preparing for lactation. After lactation the mammary gland undergoes involution (apoptosis) to return to pre-pregnancy state <sup>1</sup>.

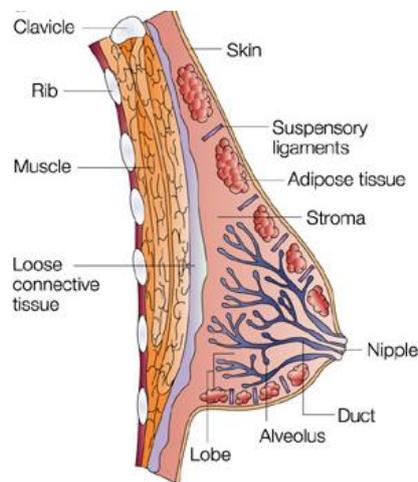


Figure1: Normal mammary gland consists of adipose tissue, stromal cells, ducts, alveoli and lobes <sup>2</sup>.

Breast tissue consists of different epithelial cells, adipocytes, blood vessels, stroma cells, fibroblast and different immune cells (figure 1). The epithelial cells from the ductal network in-between the fat cells. Basal and luminal cells are the two main types of epithelial cells in the breast tissue (figure 2). The basal epithelium consists of myoepithelial cells that form the

outer layer of the bi-layered structure of breast epithelium while the luminal cells make up the inner layer <sup>1</sup>.

### 1.1.2 Breast cancer incidence and risk factors

Breast cancer is one of the most common causes of cancer-related deaths in women <sup>3</sup>. In Norway in 2011, 3094 women and 28 men were diagnosed with breast cancer from the total of 16189 who were diagnosed with cancer <sup>4</sup>. In the same year, 612 Norwegians died from breast cancer. The 5-year survival rate in Norway between 2007-2011 were almost 89% for women, compared to 67% in 1972-1976 <sup>4</sup>. This can be explained by earlier diagnosis and better treatment, and/or more targeted treatment <sup>5</sup>.

Besides gender and age, risk factors for breast cancer include genetic susceptibility due to mutations in breast cancer 1, early onset (*BRCA1*) and breast cancer 2, early onset (*BRCA2*) gene and some other more rare predisposing genes. For sporadic cancers, radiation exposure, a previous diagnosis of a pre-invasive lesion, number of children and age when giving birth, hormone-replacement therapy and obesity may influence the risk for developing breast cancer <sup>6</sup>.

Six biological capabilities of tumors were described by Hanahan and Weinberg <sup>7</sup> as the "hallmarks of cancer". Human tumors share some common features including cell signaling, proliferation and metastasis capabilities; sustainability to proliferation, evade growth suppressing signals, enable replicative immortality, resists apoptosis, induce angiogenesis and activate invasion and metastasize <sup>7</sup>.

### 1.1.3 Development and progression of breast cancer

Carcinomas refer to tumors arising from epithelial cells <sup>8</sup>. Invasive breast cancer develops over time from pre-existing benign lesions; however, not all benign lesion develops into invasive breast cancer. Intraductal hyperplasia and atypical ductal hyperplasia are early steps in breast cancer development that later may evolve into ductal or lobular carcinomas *in situ* (DCIS and LCIS) (figure 2). At the *in situ* stage, the number of stromal cells increases and the carcinoma cells eventually break through the basal membrane and invade surrounding breast tissue. Hyperplasia and *in situ* carcinomas have malignant properties, such as uncontrolled cell proliferation, but are considered as pre-invasive because they lack the ability to invade

and metastasize. Metastasis greatly increases the patient's likelihood of dying from breast cancer, when the tumor cells spread to other organs in the body <sup>9, 10</sup>. Ductal and lobular subtypes are the most frequent tumor types in the breast, and almost 80% of preinvasive and invasive breast tumors are of ductal subtype <sup>11</sup>.

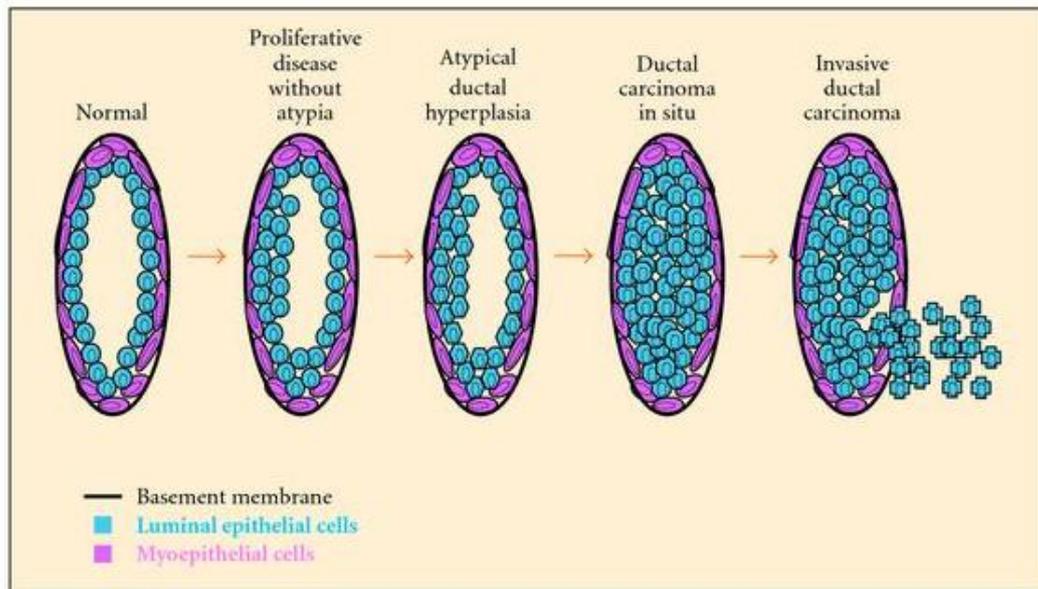


Figure 2: The hypothetical breast tumor progression from normal to hyperplasia, to *in situ*, to invasive and metastatic carcinoma stages. The normal breast ducts are composed of luminal cells in the inner layer and basal cells (myoepithelial cells) covered by a basement membrane <sup>12</sup>.

Breast cancer is a heterogeneous disease which should not be considered as just one single disease but rather a collection of diseases. Several factors contribute to the heterogeneity including a wide variability between individuals, between tumors and within tumors as well as differences at the molecular level. Intertumor heterogeneity between patients and intratumor heterogeneity within a tumor, such as the presence of both basal-like and luminal tumor cells within one single tumor contributes to the overall heterogeneity observed in breast cancer <sup>13</sup>. Two main genes are known to be involved in hereditary breast cancer, *BRCA1* and *BRCA2* which are found mutated in up to 20% of breast cancers. The remaining 80-85% of breast cancers are sporadic and caused by genetic <sup>14</sup> and epigenetic alterations. Epigenetics is a reversible process and regulates gene expression through DNA methylation, histone modification, nucleosome position, non-coding RNAs and microRNA. Failure in the epigenetic machinery, leading to activation or inhibition of signaling pathways may lead to cancer. Genetic and epigenetic alterations work together in initiation and progression of

cancer<sup>15</sup>. Mutations in tumor suppressor genes and/or in proto-oncogenes lead to cancer. Proto-oncogenes are normally involved in cell proliferation or apoptosis and tumor suppressor genes normally act to control cell cycle progression and cell division<sup>16</sup>. Phosphatase and tensin homolog (*PTEN*) and tumor protein p53 (*TP53*) are tumor suppressor genes and are often found malfunctioned in breast cancer. *TP53* is a frequently mutated gene in human breast cancers<sup>17</sup>. Mutations in several tumor suppressor genes and oncogenes have been detected in breast cancers, but a major challenge is the diversity in breast cancer which results in large genetic diversity<sup>18</sup>.

#### 1.1.4 Signaling pathways

Up-regulation of cell cycle controlling signaling pathways are associated with development and progression of tumors and are considered as mitogens that drive cells through cell division<sup>19</sup>. Aberrant Hedgehog signaling may lead to tumor development<sup>20</sup>. The Mitogen-activated protein kinase (MAPK) signaling pathway is a cell cycle controlling pathway that is associated with breast cancer<sup>21</sup>. In addition, estrogen and progesterone receptor are also involved in cellular proliferation<sup>22,23</sup>. Estrogen receptor, progesterone receptor and epidermal growth factor receptor and human epidermal growth factor 2 involved in the MAPK signaling pathway, are used to characterize breast tumors in the clinic<sup>24</sup>.

#### 1.1.5 The hedgehog signaling pathway

The Hedgehog (Hh) gene were first discovered in *Drosophila*<sup>25</sup>. The Hh name originates from Mohler works in 1988, where he discovered that *Drosophila* larvae with Hh gene mutation had short and spined phenotype that resembled the spikes of a hedgehog<sup>26</sup>. There are three Hh gene homologues, Sonic- (*SHH*), Desert- (*DHH*) and Indian hedgehog (*IHH*), which have different biological functions in mammals and controls multiple different developmental processes. The Hh signaling pathway has an important role in embryogenesis where Hh signaling controls cell fate, tissue patterning, proliferation, survival and differentiation. In adults, Hh signaling is involved in regulation of tissue homeostasis, regeneration and stem cell maintenance. The Hh pathway is activated by extracellular signal, which subsequently leads to intracellular regulation of gene expression. The Hh proteins are lipid modified and membrane anchored<sup>27-29</sup>. The transmembrane protein Dispatched (Disp) liberates anchored Hh proteins from the cholesterol and releases it from internal or surface membrane so that the

Hh protein can be secreted<sup>30</sup>. The 12 pass transmembrane protein receptor Patched (Ptch) is located at the base of primary cilium, an microtubule-based organelle (figure 3). Two homologues genes of PTCH, named PTCH1 and PTCH2 exist. The Hh ligand binding to the Ptch receptor is regulated by additional proteins and the expression of these genes helping Hh ligand to bind to Ptch is downregulated in response to Hh signaling giving rise to a negative feedback loop. In the absence of ligand Ptch inhibits the 7 pass transmembrane protein Smoothed (Smo) by keeping it in intracellular vesicles. Upon binding of extracellular Hh protein to Ptch, Ptch becomes inactive and losses the ability to inhibit Smo. Smo translocates to the cell surface in the primary cilium and induces downstream signaling. Activated Smo transduces an intracellular signal that finally activates the families of Glioma associated oncoproteins (GLIs) by removing GLI from an inhibitor multiprotein complex. There are three members of the GLI family, GLI1, GLI2 and GLI3. GLI1 is a transcription factor, GLI2 have both activation and inhibitor function, while GLI3 is a transcription inhibitor. The outcome of activated Hh signaling pathway depends on the balance between GLI activation and repressor forms. The Hh signaling pathway has several target genes, including GLI1<sup>27-29</sup>.

Aberrant Hh signaling may lead to tumor development and may be caused either by overexpression through ligand-dependent activation or by ligand-independent overexpression or mutation of the components in the Hh pathway. Tumor cells can produce Hh ligand and stimulate neighboring cells through paracrine signaling or the tumor cells can stimulate themselves through autocrine signaling<sup>29</sup>. Mutations in *SHH*<sup>31</sup>, *GLI1*<sup>32</sup>, *SMO*<sup>33</sup> and *PTCH1*<sup>34</sup> genes have been detected in different types of cancer, while others have not been able to confirm mutations in these genes<sup>35</sup>. Information concerning altered genetics and expression of Hh pathway components, in breast cancer subtypes or progression is limited. However, overexpression of GLI1 has been shown to result in developing mammary gland tumors and thereby suggesting that GLI1 can function as an oncogene<sup>20</sup>. From immunohistochemical analyzes it has been suggested that breast tumors with nuclear GLI1 expressions correlate with poorer survival<sup>36, 37</sup>. Additional evidence supporting a role for Hh signaling in breast cancer comes from high resolution CGH (comparative genomic hybridization) analysis revealing frequent loss of PTCH1 (9q22.1-q31) and amplification of GLI1 (12q13.2-q13.3) chromosomal regions in breast cancer samples and cell lines<sup>38, 39</sup>. The Hh signaling pathway is also important for normal development of the mammary gland<sup>40</sup>.

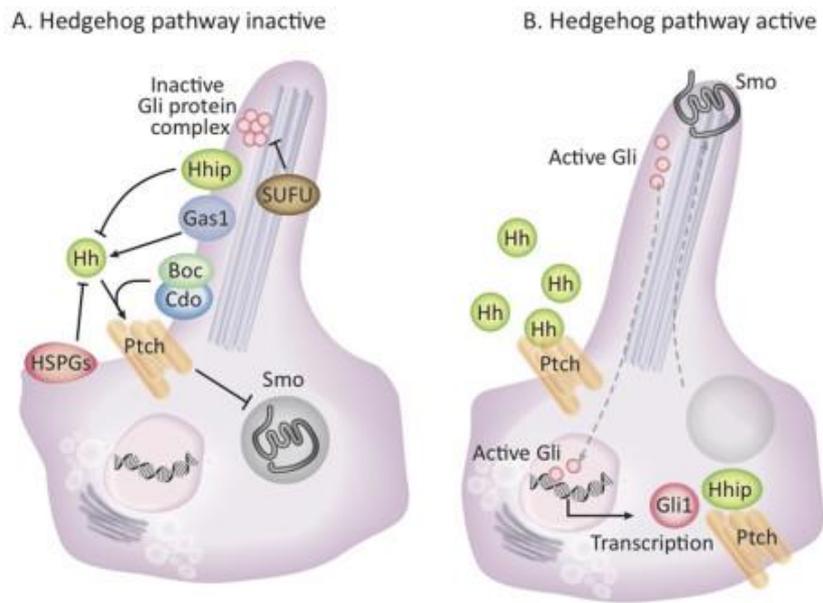


Figure 3: The Hedgehog signaling pathway when in an inactivated (A) and activated (B) state. In the presence of ligand, PTCH stops to inhibit Smo and Smo can translocate to the cell surface and releases Glioma associated oncoproteins (GLIs) from an inhibitor multiprotein complex. Activated GLI1 translocates to the nucleus and alters gene expression <sup>29</sup>.

### 1.1.6 Epidermal growth factor receptor

Epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase (RTK) and belongs to V-erb-b2 erythroblastic *leukemia* viral oncogene homolog (ERBB) family consisting of 4 family members. EGFR is also known as ERBB1 and HER1 and is a transmembrane protein crossing the cell membrane once with extracellular ligand-binding domain, and intracellular enzyme activity. Binding of signaling protein (ligand) to the ligand-binding domain, causes the receptor to dimerize with another ERBB and the receptors crossphosphorylate each other (transautophosphorylation) on multiple tyrosines to become activated. Phosphorylated tyrosines act as docking sites for downstream intracellular signaling proteins that relay the signal downstream, by the binding-protein is conformational changed, phosphorylated at tyrosine residues or by coming in contact with other proteins in the signaling pathway. Activated signaling proteins leads to assembly of intracellular signaling complexes, which transfer the signal further down the signaling pathway. Different RTKs bind different combinations of signaling proteins and EGFR activates MAPK, PI3-K and PLC- $\gamma$  signaling pathways and trigger different responses.

EGFR is involved in cell growth and proliferation<sup>41</sup> during early stages of pregnancy and during lactation in mammary gland<sup>42, 43</sup>. Over expression of *EGFR* have been reported in human breast cancer<sup>44, 45</sup> and about 45% of breast tumors have *EGFR* overexpression<sup>46</sup>. *EGFR* gene amplification have been reported but it is seldom<sup>45, 47</sup>. EGFR positive tumors have been found in basal-like and triple-negative breast tumors.<sup>48, 49</sup>. EGFR positive breast tumors are associated with poor prognosis<sup>46, 50</sup>.

### 1.1.7 Human epidermal growth factor receptor 2

Human epidermal growth factor 2 (HER2) is a member of the ERBB family and does not have its own ligand that can bind and activate the receptor and therefore heterodimerizes with other ERBB to activate downstream signaling pathways<sup>51</sup>. *HER2* are low expressed in normal breast cells<sup>52</sup> and higher expressed in proliferating stages during puberty and early pregnancy<sup>53</sup>. *HER2* overexpression have been found in breast tumors<sup>45, 54, 55</sup> and are often caused by gene amplification<sup>45, 56</sup> whereas 2% HER2 positive tumors have mutated *HER2* gene<sup>55</sup>. Amplification of *HER2* gene increase transcription and synthesis of *HER2*. *HER2* expressing breast cancer cells have the ability to proliferate, invade and block apoptosis leading to tumor growth and differentiation<sup>57</sup>. About 20% of breast tumors are HER2 positive<sup>58</sup> and HER2 positive tumors are associated with poor prognosis<sup>45, 57</sup>.

### 1.1.8 MAP-Kinase signaling pathway

Mitogen-activated protein kinase (MAPK) signaling pathway is a cascade of 4 kinases. The adaptor protein Grb binds to phosphorylated EGFR (figure 4). Ras guanine nucleotide exchange factor (RasGEF) (a.k.a. Sos) binds to Grb and can activate several downstream signaling pathways including the MAPK pathway. Ras is a GTPase anchored to the plasmamembrane and inactive when bound to GDP, GDP is removed by Sos so GTP can bind and activate Ras. Tyrosine phosphorylation at EGFR and active Ras is short lived and relaying signal downstream extends the signal. Active Ras phosphorylates MAP kinase kinase kinase (MAPKKK) (a.k.a. Raf) that phosphorylates MAP kinase kinase (MAPKK) (a.k.a. Mek) that in turn phosphorylates MAP kinase (MAPK) (a.k.a. Erk). Erk translocates to the nucleus where it phosphorylates proteins involved in gene expression of cell proliferation and cell differentiation genes.

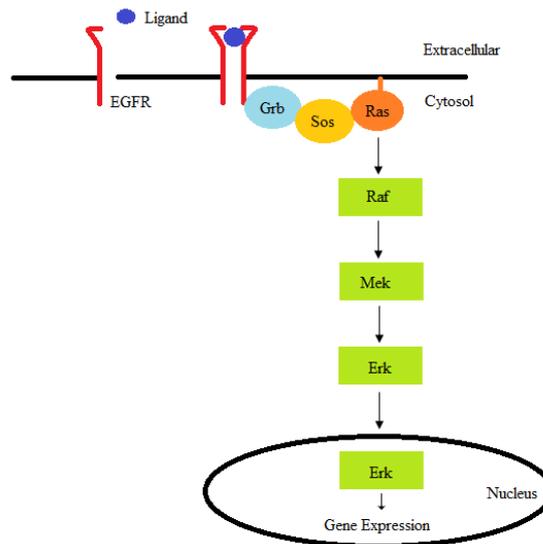


Figure 4: The MAPK signaling pathway: Adaptor protein Grb binds to phosphorylated tyrosine at ligand bound activated EGFR. Sos binds to Grb and activates Ras that activates the Raf, Mek, Erk phosphorylation cascade and Erk translocates to the nucleus and alters gene expressions.

H-, K- and N-Ras are Ras family members and part of the MAPK signaling pathway. *RAS* has been reported to be overexpressed in breast tumors compared to normal breast tissue. There have not been large studies to say how frequent *RAS* is overexpressed in breast tumors but about 50% of breast tumors have *RAS* overexpression in a small study<sup>59</sup>. *RAS* overexpression can be due to aberrant activation of the upstream signaling pathway or due to *RAS* mutations. Mutated *RAS* is rare and found in less than 5% of breast tumors<sup>60, 61</sup>. Mutations or overexpression of RTKs, including *HER2* and *EGFR* can lead to overactivation of Ras and upregulation of MAPK signaling pathway<sup>62</sup>.

MAPK comprises several Erk variant, including Erk 1 and 2, which have similar function so they are collectively termed Erk1/2. Phosphorylated Erk1/2 (p-Erk1/2) is the active form of the protein<sup>63</sup> and is involved in cell proliferation<sup>64</sup>, cell growth in normal cells<sup>65</sup> and are related to cancer cell progression. Erk1/2 activation can be due to mutations and up-regulation of other factors in the MAPK signaling pathway. Erk activation also up-regulates EGFR expression promoting a positive loop for tumor growth<sup>62</sup>. Breast cancer show lower p-ERK1/2 than normal breast cells<sup>63</sup>. p-ERK1/2 over expression is found in breast cancer<sup>66</sup>. p-Erk1/2 positive breast cancer and activation of the MAPK signaling pathway is associated with poor prognosis<sup>63, 67</sup>.

### 1.1.9 Estrogen receptor and progesterone receptor

Estrogen receptor (ER) and progesterone receptor (PR) are steroid hormone receptors and are closely linked; 96% of steroid positive cells express both ER and PR<sup>68</sup>. The nuclear receptors ER and PR are ligand dependent and estradiol binds to ER<sup>69</sup> and progesterone binds to PR<sup>70</sup>. In normal mammary gland ER and PR are activated by paracrine signaling<sup>71</sup>. ER and PR are involved in cell proliferation in breast tissue<sup>22,23</sup>.

ER is a nuclear receptor<sup>72</sup> which stimulates cell proliferation<sup>23</sup>. ER has two isoforms, ER-alpha and ER-beta. ER-alpha is a transcription activator and is inhibited by ER-beta and expression level of both ERs is important in cellular responses<sup>73</sup>. The ER isoforms have distinct distribution pattern in normal mammary gland where ER-alpha is found to be expressed in luminal cells and ER-beta in basal cells<sup>74</sup>. ER-alpha is heterogeneously expressed in mammary gland<sup>68</sup> and necessary for normal breast development<sup>75</sup> and is involved in ductal proliferation during development and pregnancy<sup>76</sup>. ER-alpha has been reported up-regulated in ER positive breast cancers<sup>77,78</sup>.

PR is primarily found in the nucleus but can also be located in small amount in the cytoplasm<sup>79</sup>. There are two isoforms of PR, A and B which have their own promoters<sup>80</sup>. PR-B is the activator for progesterone target genes, while PR-A inhibit PR-B activity<sup>81</sup>. PR functions as a transcriptional factor<sup>22</sup> and is expressed in normal breast luminal cells but not in all luminal cells<sup>82</sup>. PR is involved in development of the ductal epithelium in the mammary glands<sup>22</sup> but is not required for normal epithelium development in all cells<sup>83</sup>. PR is found in early pregnancy<sup>82</sup> and involved in lactation when the lobular-alveolar system proliferates<sup>22</sup>.

In normal breast cells ER or PR positive cells are negative for the proliferation marker Ki-67 meaning that ER and PR containing cells do not divide under normal conditions<sup>68</sup> and use paracrine signaling<sup>71</sup>. In breast tumors on the other hand the, ER or PR positive cells show proliferation<sup>68, 82</sup>. Since PR and ER positive cells normally show no proliferation, this suggests that the signaling switches from paracrine to autocrine and this contributes to tumorigenesis<sup>71, 84</sup>. A paracrine feedback loop from receptor positive cells probably control proliferation of stem cells<sup>68</sup>.

ER positive tumors are found in about 80% of breast cancer. 65% is of breast cancer of ER positive tumors are also PR positive. 25% of breast cancer is negative for both receptors<sup>85</sup>. Breast tumors with both ER and PR positive tumors, have higher survival rate than only PR or

ER positive tumors or double negative breast tumors. The double negative tumors have the highest mortality rate<sup>86, 87</sup>.

### 1.1.10 Cytokeratins

Cytokeratins are keratin-containing intermediate filament proteins and are part of the cytoskeleton in epithelial cells. The intermediate filaments can be found in cytoplasm, between cells, and just beneath the nuclear membrane and are involved in desmosomes (cell-cell contacts) and hemidesmosomes (cell-matrix contacts). The intermediate filaments thereby give mechanical strength by preventing bending and twisting of epithelial cells and epithelial tissue. Cytokeratins are divided in type I that are acidic and type II that are basic. The keratins form heterodimer filament consisting of one type I and one type II keratin. Two heterodimers form a tetramer, and eight tetramers form the intermediate filament.<sup>88</sup> The cytokeratins are numbered according to their molecular weight and isoelectric point<sup>89</sup> and there are at least 54 functional cytokeratins. Cytokeratins are also called keratins according to the new nomenclature published in 2006, but the old nomenclature is still in use<sup>90</sup>.

In 1979 Franke et al. described that human epithelial cells have different intermediated-sized filament that can be distinguished by immunological methods<sup>91</sup>. Cytokeratins are expressed in different epithelial cells and can thereby be characterized by their cytokeratin pattern. Cytokeratins are useful in diagnosing epithelial cancer (carcinomas) which can reveal the epithelial tissue where the cancer originated<sup>89, 90, 92</sup>. Cell type heterogeneity in epithelial tissues might explain why a cytokeratin dominates the tumor but not the entire tissue. In mammary gland tissue there are three different epithelial cells (myoepithelial, ductal and secretory cells) and which have their one distinct cytokeratins characteristics<sup>89, 93</sup>. The diversity of cytokeratins in epithelial cells is useful in determine if the breast tumor is of luminal or basal origin<sup>94</sup>.

Cytokeratin 5 (K5) is a marker for basal epithelial and progenitor cells in the normal breast. In other tissues K5 only stain basal cells<sup>95</sup>. Staining of K5 in cancer cells and in both luminal and basal cells in normal breast have been reported<sup>96-99</sup> which can be explained that K5 stains progenitor cell in breast tissue that have stem-cell properties and can develop to both glandular epithelial and myoepithelial cells<sup>98, 100</sup>.

Cytokeratin 6 (K6) is a basal keratin<sup>101</sup> and expressed during mammary gland development and sporadic expressed in mature gland<sup>102</sup>. K6 is shown to stain breast cancer tissue<sup>103-105</sup>.

Cytokeratin 18 (K18) was found to stain mammary gland back in 1982<sup>106</sup>. K18 stains luminal cells in normal mammary tissue<sup>92</sup> but do not stain basal cells, myoepithelial cells or intratumoral lymphocytes<sup>107</sup>. K18 is shown to stain mammary gland tumors<sup>108, 109</sup>.

#### 1.1.11 Ki-67

Ki-67 antibody recognizes the nuclear Ki-67 protein that is associated with proliferating cells and can be used in immunohistochemistry method to detect tissue with proliferation and in risk of malignant transformation. Ki-67 is a good protein for detecting cell growth and all proliferation activity since it detects all active phases of the cell cycle except the resting stage G<sub>0</sub>. The Ki-67 is probably involved in transcription of ribosomal RNA<sup>110</sup>. Ki-67 staining is not found in normal breast tissue<sup>97</sup> but has been shown to stain proliferating cancerous breast cells<sup>111-113</sup>. Higher expression of Ki-67 is associated with lower diseases-free survival and overall survival<sup>114</sup>.

#### 1.1.12 Breast cancer subtypes

Breast tumors can be classified in to four molecular subtypes based on gene expression patterns<sup>115</sup>. As an extension to this study, two additional studies refined the classification to include five different subgroups; luminal A, luminal B, ERBB2+, basal-like and normal breast-like subtypes, which were validated in independent patient cohorts<sup>116, 117</sup>. The ERBB2+ subgroup show overexpression and/or amplification of *ERBB2* (also known as *HER2*) and no expression of ER or PR. Luminal A and luminal B tumors show expression of ER and/or PR and are distinguished from each other by expression of other genes such as proliferation associated genes which are high in luminal B<sup>118, 119</sup>. The basal-like subgroup separates form the luminal groups by having no ER-alpha expression, high expression of cell cycle genes and express basal keratins K5/6 and K17. *BRCA1* mutation carriers often develop basal-like tumors. The normal like subgroup has low expression of luminal epithelial genes, high expression of basal epithelial genes and high expression of genes originating from adipose tissue and non-epithelial cell types<sup>119</sup>. This subtype is often omitted in clinical studies (figure 5). In later years a new subgroup have been identified; claudin-low which is a subgroup with low expression of genes involved in tight junctions and cell-cell adhesion<sup>120</sup>, but which is not commonly used for breast cancer classification<sup>121</sup>. The different subgroups are associated with different prognosis where basal-like and ERBB2+ subtypes have the poorest prognosis<sup>119</sup>. As a surrogate classification, the molecular subtypes can be classified

based on immunohistochemistry using PR, ER, HER2 and Ki-67 (figure 5) <sup>122</sup>. Breast tumors that are negative for ER, PR and HER2 are also referred to as triple-negative, comprise 12-17 % of breast cancers and partly overlap with the basal-like subgroup.

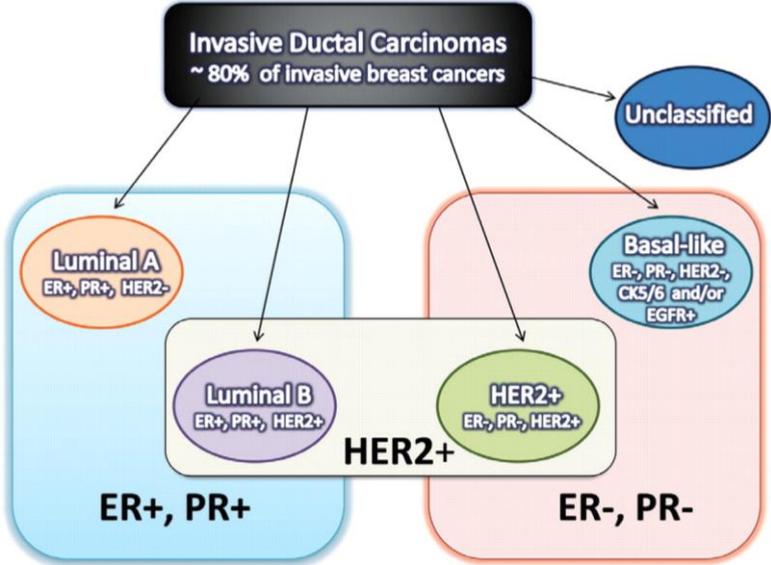


Figure 5: Breast cancer can be divided in 4 subtypes based on IHC markers. Luminal A, luminal B, HER2+, basal-like and in addition there will be unclassified tumors. The different subtypes are associated with different gene expression pattern of ER, PR, HER, EGFR and CK5/6 <sup>24</sup>.

### 1.1.13 Functional properties of mutated genes

Table 1: The main function and involvement in signaling pathway of the 20 genes sequenced with Sanger sequencing technology in this thesis. The information is gathered from National Center for Biotechnology Information gene database.

<b>Gene Name</b>	<b>Full gene name</b>	<b>Gene Function/Signaling pathway</b>
<b>ATP13A3</b>	ATPase type 13A3	Transports cations across membranes
<b>BRCA2</b>	Breast cancer 2, early onset	DNA repair, tumor suppressor gene
<b>BRCC3</b>	BRCA1/BRCA2-containing complex, subunit 3	DNA damage response
<b>BTNL6</b>	Butyrophilin-like 6	Not known
<b>CDK4</b>	Cyclin-dependent kinase 4	Important for cell cycle G1 phase progression
<b>COMMD7</b>	COMM domain containing 7	Involved in negative regulation of NF-kappaB transcription factor activity negative regulation of transcription, tumor necrosis factor-mediated signaling pathway
<b>FAM179B</b>	Family with sequence similarity 179, member B	Not known
<b>FOXP1</b>	Forkhead box P1	Tumor suppressor, regulates tissue and cell type-specific gene transcription
<b>IL7R</b>	Interleukin 7 receptor	Plays a critical role in the V(D)J recombination during lymphocyte development and involved in PI3K-Akt signaling pathway
<b>KDM6A</b>	Lysine (K)-specific demethylase 6A	Wnt signaling pathway, and histone methylation
<b>KRAS</b>	Kirsten rat sarcoma viral oncogene homolog	Oncogene, involved in the MAPK signaling pathway
<b>LPIN1</b>	Lipin 1	Involved in lipid metabolism
<b>MAML2</b>	Mastermind-like 2 (Drosophila)	Transcription activator involved in Notch signaling pathway
<b>MYOM1</b>	Myomesin 1	Interconnects the major structure of sarcomeres
<b>OVGP1</b>	Oviductal glycoprotein 1	Involved in fertilization and reproduction
<b>PSIP1</b>	PC4 and SFRS1 interacting protein 1	Involved in heterochromatin and gene transcription
<b>PTCH1</b>	Patched 1	Receptor for Hedgehog proteins
<b>SEMA6D</b>	Sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D	Involved in axon signaling
<b>SFI1</b>	Sfi1 homolog, spindle assembly associated (yeast)	Cell cycle, G2/M transition of mitotic cell cycle
<b>STAT6</b>	Signal transducer and activator of transcription 6, interleukin-4 induced	Transcription factor, a central role in exerting IL4 mediated biological responses
<b>TSC1</b>	Tuberous sclerosis 1	Involved in PI3K-Akt signaling pathway and, mTOR signaling pathway

### 1.1.14 Model systems

Mice are excellent model organism when researching human diseases. The anatomy, physiology and genetics are very similar in mice and humans. More than 95% of the mouse genome is similar to the human genome, making mice excellent model organism when studying human diseases caused by mutations and diseases when multiple genes are involved.

Another advantage is that mice have short generation time and fast development reducing time used for breeding, preparing and conducting experiments <sup>120, 123, 124</sup>.

### 1.1.15 Ethics

The ethical aspect of using mice in research is conflicted. In Norway the animal welfare law protects the animal from being exposed to unnecessary suffering but that it is allowed to expose the animal suffering as long as it is not unnecessary suffering <sup>125</sup>. Unnecessary suffering means that the animal must be treated with respect, and not be treated irresponsibly, brutal, vicious, exposed of unnecessary suffering, victim of animal abuse and the animals animal instincts and natural needs shall be protected. The term unnecessary suffering gives the law flexibility and is a judgment call if the suffering is justified <sup>126</sup>. Use of animal research in Norway has to be approved by the animal research committee <sup>127</sup>. Use of transgenic mouse in this project was approved by Stockholm South Animal Ethics committee.

## 1.2 Methodological background

This chapter describes the methodological background for the methods used in this thesis.

### 1.2.1 Mice

Research with mice began in the early 1900s <sup>128</sup>. Human and mouse are very similar genetic and physiologic <sup>123</sup> e.g. GLI1 shows 85% amino acid homology between in human and mouse <sup>40</sup>. Mice are therefore a good and relevant model organism for studying human physiology and pathology, including GLI1 induced mammary gland tumors. There is no single mouse models that can be used to study all human diseases. Thus, different genetically modified mouse models are used to address different conditions studies <sup>129</sup>. C57Black/6 (C57BL/6) is one of the most used strains and is a black mouse with low incident of spontaneous mammary gland tumors, which makes it suitable to induce mammary tumors (figure 6) <sup>130</sup>. The FVB/N mouse strain has high reproductive activity, which is an advantage in research with transgenic mice and has been shown to be more prone to develop mammary gland tumors in transgenic mouse models <sup>131, 132</sup>. Non-obese diabetic/severe combined immunodeficient (NOD/SCID) mouse is white, has reduced immunsystem and can be used for xenograft transplantation of non-syngenic tissue, including human tissue (figure 7) <sup>133</sup>.



Figure. 6: C57BL/6 mouse. Picture obtained from: [http://jaxmice.jax.org/images/jaxmicedb/featuredImage/005304\\_lg.jpg](http://jaxmice.jax.org/images/jaxmicedb/featuredImage/005304_lg.jpg)



Figure 7: NOD/SCID mouse. Picture obtained from: [http://jaxmice.jax.org/images/jaxmicedb/featuredImage/005557\\_lg.jpg](http://jaxmice.jax.org/images/jaxmicedb/featuredImage/005557_lg.jpg)

In conditional inducible transgenic mice the tetracycline-controlled transcriptional activation (tTA) and reverse tetracycline-controlled transcriptional activator (rtTA) systems are used to control gene transcription by turning on and off transcription by administration of tetracycline antibiotics<sup>134</sup>. Mouse mammary tumor virus (MMTV) is used to drive expression of transgenes in mammary gland. Insertion of tetracycline responsive elements (TRE) in front of the transgenic gene of interest is required to respond to tetracycline induced expression. This system was used to direct expression of human GLI1 to the mouse mammary gland in double transgenic mice, MMTVrtTA;TREGLI1, obtained from crossing heterozygote single transgenic mice carrying the TREGLI1 and MMTVrtTA, respectively<sup>40</sup>.

### 1.2.2 Hematoxylin and eosin staining

Hematoxylin and Eosin (H&E) staining was first described as in 1877<sup>135</sup> and is used to stain cellular components in fixed and sectioned tissue, and to identify and characterize cancer. Hematoxylin is a positively charged basic dye that binds to negatively charged molecules, for example DNA in nucleus and colors them purple-blue. Eosin is a negatively charged acid dye that counter stains basic structures pink, such as cytoplasm, and intracellular membranes.

### 1.2.3 Immunohistochemistry

Immunohistochemistry (IHC) is an antibody based technique first published by Albert Coons in 1941<sup>136</sup> and can be used to identify tissue components, study localization and distribution of proteins.

### 1.2.3.1 Antibody-antigen binding

Immunoglobulins (Igs) are produced and secreted by activated B-cell lymphocyte in an immune response. Igs are called antibodies and binds to foreign antigen, which will result in activation of the immune system and hinder further invasion and survival of the microbe.

Antibodies are glycoproteins that consist of four polypeptide chains, two light and two heavy chains (figure 8). Antibodies have two identical halves with the same antigen-binding site that are held together by non-covalent and covalent (disulfide) bonds. Antibodies have a variable region giving diversity to the antigen binding sites and a constant region also called fragment crystallizable (Fc) region responsible for the functional properties, including activation the of complementary system or binding site for phagocytic cell. The N-terminal ends of the light and heavy chains come together to form the antigen binding site in fragment antigen binding (Fab) region that binds to the antigen. The hinge region is flexible and the distance between antigen binding sites can vary and this increases the efficiency of antigen binding. Antigenic determinants (epitopes) are the antigen binding site for antibodies and the antigen may have multiple epitopes. An antigen can be defined as everything that can be recognized by lymphocytes antigen receptors. The binding of antibody to antigen is a reversible interaction and the binding strength depends on the complementary fit between antibody and antigen.

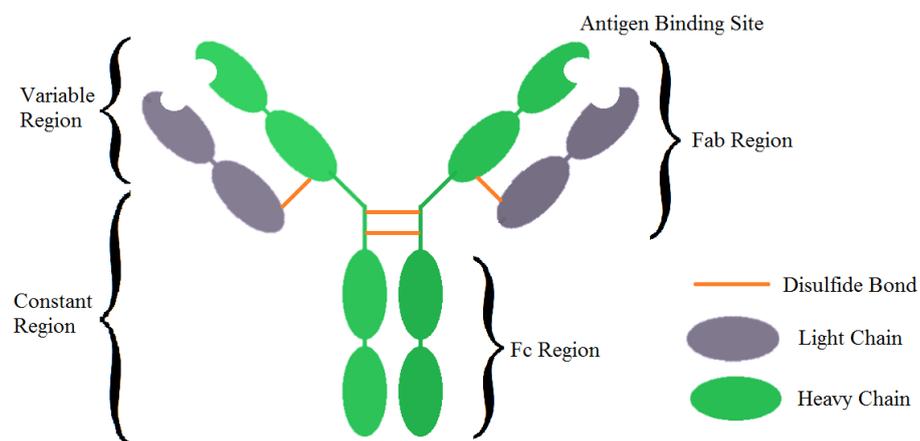


Figure 8: The drawing shows the structure of antibodies, with the light and heavy chains held together by disulfide bonds and the antigen binding site in the fragment antigen binding (Fab) region. The chains have variable regions that give the diversity of the antigen binding sites. The fragment crystallizable (Fc) region in the constant region determines the mechanism used to destroy the targeted antigen.

In an immune response a single B cell can be activated giving a monoclonal response with only one identical antigen binding property (figure 9). The monoclonal response can produce different classes of Igs (IgM, IgG, IgA, IgE and IgD) that have different number of antigen binding sites, but have the same antigen binding property but have different functions. The immune response could also be of polyclonal when many different lymphocytes are activated. Polyclonal antibody-antigen binding is the most common activation.

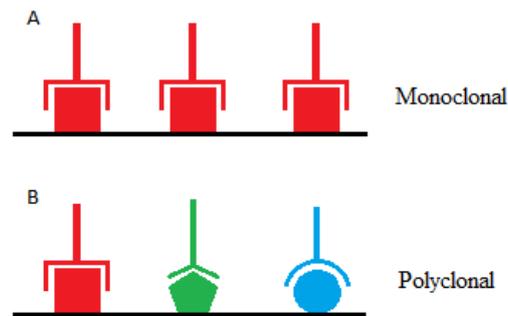


Figure 9: Schematic illustration of monoclonal and polyclonal antibodies. A: Monoclonal antibodies are specific to one epitope type. B: Polyclonal antibodies include different types of antibodies and can bind to several epitopes.

#### 1.2.3.2 Avidin-Biotin complex, immunohistochemistry, and indirect method

Fixation of tissue preserves the tissue architecture, cell morphology and antigen properties. The fixed tissue is embedded in paraffin to make a solid block that is easy to cut the tissue into a few  $\mu\text{m}$  thick slices. The water containing tissue is not soluble with paraffin and removing of water through a dehydration process with alcohol and xylene clearing replaces the water in the tissue. The paraffin in the sectioned tissue on glass slide is removed through deparaffinization, otherwise the paraffin could otherwise affect the staining of the tissue. Most of the solution used in the staining procedure are aqueous does not work with dehydrated tissue. Rehydration replaces the xylene in the tissue with water required to obtain proper staining. Unlabeled primary antibodies bind to the protein's binding sites (epitope) in the tissue (figure 10). Biotinylated secondary antibodies bind to the Fc region of the primary antibodies. Primary and secondary antibodies are species specific and the secondary antibody is anti-species of the host of the primary antibody, e.g. a polyclonal primary antibody produced in a species different for rabbit. This method, unlabeled primary antibody and labeled secondary antibody is called an indirect IHC method and more sensitive than a direct method that uses a labeled primary antibody. The avidin-biotin complex method uses

streptavidin peroxidase that forms a complex by binding strongly to biotin conjugated to secondary antibody. Streptavidin peroxidase oxidizes the 3,3'-Diaminobenzidine (DAB) chromogen which results in a water-insoluble brown precipitate seen as brown color under light microscopy<sup>137</sup>. Hematoxylin counter stains the nucleus purple-blue by binding to DNA, RNA and other negatively charge molecules in the cell. Tissue cells are visualized under a light microscope after mounting of the glass slide that may require dehydration in alcohol and xylen clearing if a nonaqueous mounting media is used.

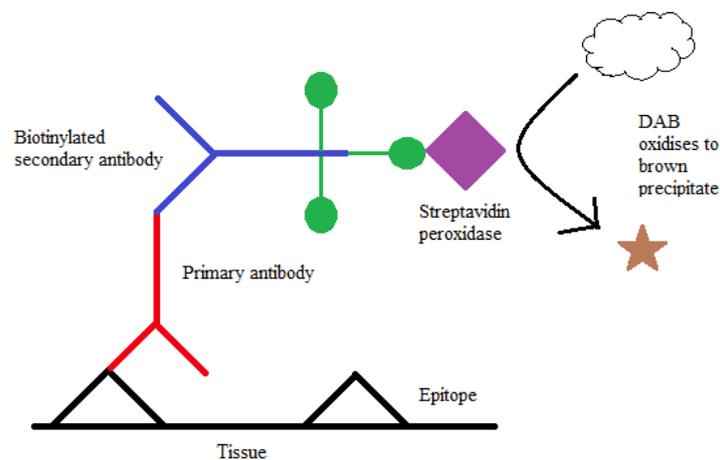


Figure 10: Indirect, Avidin-Biotin complex immunohistochemistry method where unlabeled primary antibody binds to a specific epitope. A biotinylated secondary antibody binds to the primary antibody and streptavidin peroxidase binds strongly to the biotin. DAB substrate is oxidizes and gives a brown color visible under a light microscopy.

#### 1.2.4 Chain termination sequencing (Sanger sequencing)

Sanger sequencing is a chain termination DNA sequencing method invented by Frederick Sanger in 1977<sup>138</sup>. Polymerase chain reaction (PCR) catalyzed by DNA polymerase makes a copy of a selected DNA region. Next, the double stranded (ds) DNA helix is denatured by heating to remove hydrogen bonds between the polypeptides to obtain single stranded DNA. Oligonucleotide primers are annealed to the single stranded DNA upon lowering the reaction temperature. At an optimal temperature for the DNA polymerase, the synthesis of DNA start by DNA polymerase mediated incorporation of deoxyribonucleotide triphosphates (dNTPs: dATP, dCTP, dGTP and dTTP) complementary to the DNA template. These steps of breaking hydrogen bonds, primer annealing and dNTPs incorporation, are carried out multiple times to make many copies of the targeted DNA sequence. Chain termination sequencing PCR is the second step in the sequencing procedure, where the DNA polymerase incorporates dNTPs or

one of four dideoxynucleotide triphosphates (ddNTPs (ddATP, ddCTP, ddGTP and ddTTP)) labeled with different fluorochromes (figure 12). The ddNTPs lacks the hydroxyl group at 3'-carbon required for connection with the next nucleotide (figure 13). The incorporation of a ddNTP prevents elongation and terminates the synthesis. The ddNTPs are in a small amount compared to the dNTPs, which means that DNA polymerase can synthesize for hundreds of nucleotides before a ddNTP is incorporated and the reaction is terminated. The incorporation of ddNTPs will result in PCR products of different lengths. A capillary gel electrophoresis separates the DNA molecules according to their length, with one nucleotide length resolution. A fluorescence detector recognizes and distinguishes the differently labeled ddNTPs of each fragment and thereby determines the DNA sequence. Sequencing analysis software analyzes the raw data obtained from the detector.

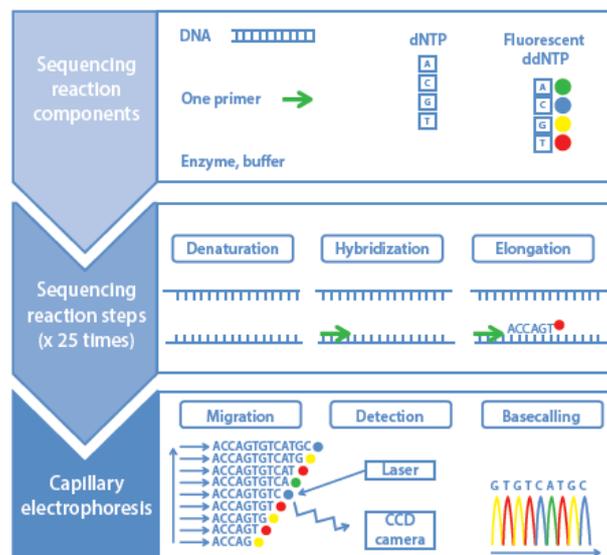


Figure 12: Sanger sequencing workflow of chain termination PCR and capillary electrophoresis; primer annealing to the DNA template, incorporation of dNTP and fluorescent labeled ddNTP, capillary electrophoresis where the sequences are separated based on the length in a gel, and detection. Figure adopted from:

<http://tools.lifetechnologies.com/content/sfs/brochures/brochure-ab-genetic-analyzers.pdf>

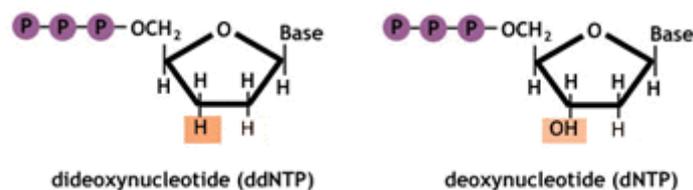


Fig 13: Illustration of the 3' difference between dNTP which extends the DNA synthesis and ddNTP which lack the OH-group at the 3' end leading to synthesis termination. Picture obtained from:

<http://missinglink.ucsf.edu/lm/molecularmethods/ddntp.htm>

### 1.2.5 Massively parallel sequencing (SOLiD sequencing)

Sequencing by oligo ligation detection (SOLiD) technology is a massively parallel next generation sequencing technology using ligation based sequencing, emulsion PCR and fluorescently labeled di-base probes (figure 11). First DNA is fragmented and is amplified by emulsion PCR where individual DNA molecules bind to small magnetic bead and amplification by producing identical fragment around one bead. The amplified templates are 3' modified so they can bind to a glass slide. The ligase mediated sequencing begins by annealing a universal primer to the adaptor sequence at the 3' end of each amplified fragment on the glass slide. Four fluorescent labeled di-bases are used as probes. 16 different probes are used in this technology where 4 different probes share the same fluorescent marker giving a two-based encoding probe. It is the 1<sup>st</sup> and 2<sup>nd</sup> base on the DNA template that decides which probe is attached. DNA ligase enzyme attach one of the four probes to the DNA template and the fluorescent signal is measured. The extension product of 3 bases for the 8mers probe is cleaved off with the fluorescent group so another probe can be attached to the template. The read length of each template is about 25-35 bases long. Five additional rounds with primer complementary to n-1 in each round are conducted to read every base in the DNA template. The sequencing can be performed in parallel with many beads attached to the glass slide <sup>139</sup>. <sup>140</sup>. The sequencing is done in millions of parallel reactions. The massively parallel sequencing gives many short reads which are assembled using reference sequence as scaffold or de novo in absent of a reference genome. The assembly gives the entire sequence of the genome of the sequenced DNA sample <sup>141</sup>. The depth of the exome sequencing obtained from this massively parallel sequencing is critical for the identification of mutations, sequence variants or massively rearrange regions <sup>142</sup>.

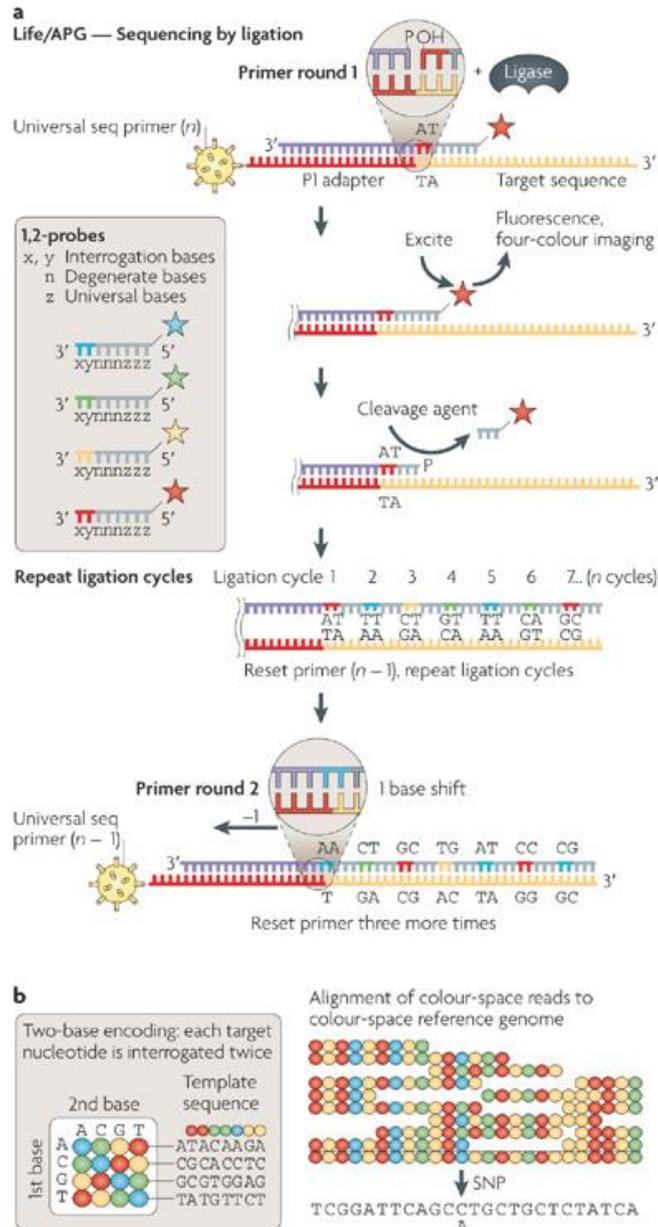


Figure 11: SOLiD sequencing, massively parallel sequencing technology with ligase-mediated sequencing and 16 different probes marked with four different fluorescence labels is used for deep sequencing of genome and may reveal point mutations upon assembly<sup>143</sup>.

### 1.2.6 cDNA synthesis

Complementary DNA (cDNA) is generated by copying mRNA molecule in a reaction catalyzed by the reverse transcriptase (RT) enzyme using mRNA molecule as template. Oligo(dt) is a short single-stranded sequence of deoxy-thymine nucleotides that binds complementary to the poly(A) tail of mRNA and acts as a primer for the RT enzyme (figure 14). RT catalyzes the cDNA synthesis by incorporating deoxyribonucleotides (dNTPs) on the

RNA template. Copping of single stranded RNA chains to complementary DNA chains, creates a DNA/RNA hybrid further used as template in PCR amplification.

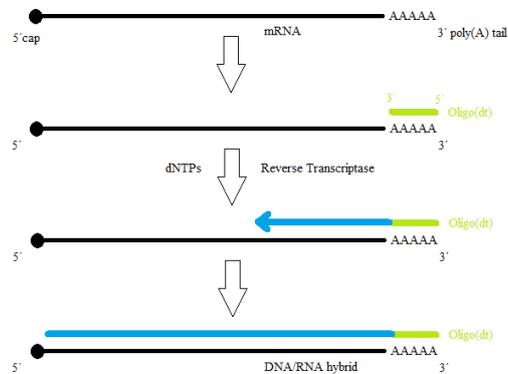


Figure 14: Synthesizing single stranded RNA to complementary DNA (cDNA) involves oligo(dt) annealing to poly(A) tail, act as a primer for the cDNA synthesis. Reverse transcriptase incorporates dNTPs and creates a hybrid of DNA/RNA in the first chain reaction.

### 1.2.7 NanoDrop 1000 spectrophotometer

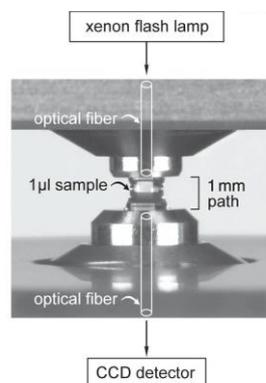


Figure 15: Nanodrop measures nucleic acids concentrations between two optical fibers using light from a xenon flash lamp detected by spectrometer. Picture obtained from: <http://nanodrop.com/Library/CPMB-1st.pdf>.

NanoDrop measures nucleic acid concentration by using the Beer-Lambert equation  $c = (A * \epsilon) / b$ .  $c$  is the concentration in  $\text{ng}/\mu\text{l}$ ,  $A$  is the absorbance,  $\epsilon$  is wavelength and  $b$  is the path length in cm defined by the distance between the two optical fibers (figure 15). Two different path lengths are used to measure the concentration. The xenon flash lamp sends light from the top optical fiber, through the sample liquid column, and the emitted light is detected by an internal spectrometer in the bottom optical fiber <sup>144</sup>.

### 1.2.8 Qubit 2.0 fluorometer

Qubit 2.0 fluorometer determines the ds DNA. A fluorophore binds to ds DNA and becomes fluorescent. The signal is measured by a photodiode in Qubit fluorometer. The fluorescent signal increases with increasing ds DNA concentration and sample concentration is calculated using the equation: Sample concentration = QF value  $\times \left(\frac{200}{x}\right)$ . The QF value is the fluorescent signal, and x is the amount sample ( $\mu\text{l}$ ) added in the tube <sup>145-147</sup>.

## 2.0 Materials

This chapter list the materials used in this thesis. The material list only includes the materials I have used, not material used by others in the group, except for the mice that have a substantial roll in this thesis and the results.

### 2.1 Mice

In this thesis, normal and tumor tissue from MMTVrtTA;TREGLI1 conditional transgenic mice were analyzed. The tissue samples have been harvested and stored previously. Primary tumor tissue was transplanted to NOD/SCID mice to maintain tumor lines in mice. Serial transplantations were carried out for more than 10 generations.

Table 2: Overview of GLI1 induced breast tumors samples analyzed with the different methods in this thesis; whole exome sequencing, Sanger sequencing of either cDNA or gDNA and immunohistochemistry. Information includes the samples tissue type, series, generation of the samples, strain of the mice in addition to the genotype of the mouse harboring the tumor. MMTVrtTA is shortened to rtTA, TREGLI1 to TRE, Lgr5-EGFR-CreERT2 to Cre and p53 indicates heterozygosity for p53 in the table.

Sample	Tissue type	Series	Generation	Genotype	Exome sequencing	Expression sequencing	Genomic sequencing	IHC
1	Breast tumor	A	Primary	rtTA;TRE;p53	+	+		+
2	Breast tumor	A	0 <sup>th</sup>	NOD/SCID				+
3	Breast tumor	A	5 <sup>th</sup>	NOD/SCID				+
4	Breast tumor	A	10 <sup>th</sup> (Late)	NOD/SCID	+	+		+
5	Breast tumor	B	Primary	rtTA;TRE;p53	+	+		+
6	Breast tumor	B	0 <sup>th</sup>	NOD/SCID				+
7	Breast tumor	B	5 <sup>th</sup>	NOD/SCID				+
8	Breast tumor	B	10 <sup>th</sup> (Late)	NOD/SCID	+	+		+
9	Breast tumor	C	Primary	rtTA;TRE;Cre;p53	+	+	+	+
10	Breast tumor	C	0 <sup>th</sup>	NOD/SCID			+	+
11	Breast tumor	C	5 <sup>th</sup>	NOD/SCID			+	+
12	Breast tumor	C	10 <sup>th</sup> (Late)	NOD/SCID	+	+	+	+
13	Breast tumor	D	Primary	rtTA;TRE	+	+		+
14	Breast tumor	D	0 <sup>th</sup>	NOD/SCID				+
15	Breast tumor	D	5 <sup>th</sup>	NOD/SCID				+
16	Breast tumor	D	10 <sup>th</sup> (Late)	NOD/SCID	+	+		+
17	Normal breast	-	15 weeks	NOD/SCID		+	+	
18	Normal breast		-	C57BL/6 FVB/N strain		+		
19	Liver	A	Primary	rtTA;TRE;p53	+			
20	Liver	B	Primary	rtTA;TRE;p53	+			
21	Liver	C	Primary	rtTA;TRE;Cre;p53	+		+	
22	Liver	D	Primary	rtTA;TRE	+			

## 2.2 Reagents and chemicals

Chemical/Reagent	Catalog number	Supplier/Manufacture
10x PBS	-	Department of Immunology, Institute for Cancer Research, Oslo, Norway
1N HCl	-	Department of Immunology, Institute for Cancer Research, Oslo, Norway
3730 Running Buffer (10X)	4335613	Applied Biosystems, Foster City, CA, USA
Acetic acid (glacial)	100063	Merck KGaA, Darmstadt, Germany
Bovine Serum Albumin	A9418-10G	Sigma-Aldrich, St. Louis, Missouri, USA
Bromophenol blue	161-0404	Bio-Rad Laboratories, Hemel Hempstead, Herts, UK
Certified PCR Agarose	161-3105	Bio-Rad Laboratories, Hemel Hempstead, Herts, UK
Deoxynucleotide triphosphates ATP	4026	Takara Biotechnology Shiga, Japan
Deoxynucleotide triphosphates CTP	4028	Takara Biotechnology Shiga, Japan
Deoxynucleotide triphosphates GTP	4027	Takara Biotechnology Shiga, Japan
Deoxynucleotide triphosphates TTP	4029	Takara Biotechnology Shiga, Japan
EDTA disodium salt	100935V	BDH, VWR International, Radnor, PA, USA
Etanol Absolutt Prima	-	Kemetyl Norge AS, Vestby, Norway
Etanol rektifisert	-	Kemetyl Norge AS, Vestby, Norway
Ficoll PM 400	F4375	Sigma-Aldrich, St. Louis, Missouri, USA
GelRed Nucleic Acid Gel Stain	41003-1	Biotium, Hayward, CA
Goat serum	G9023	Sigma-Aldrich, St. Louis, Missouri, USA
Hydrogen peroxide solution	216763	Sigma-Aldrich, St. Louis, Missouri, USA
MAYER HTX Staining Solutions	01820	Histolab, Gothenburg, Sweden
Methanol	106009	Merck KGaA, Darmstadt, Germany
Oligo(dT) <sub>20</sub> Primer	18418-020	Invitrogen, Carlsbad, CA, USA
o-Xylene for synthesis	808697	Merck KGaA, Darmstadt, Germany
Pertex mounting medium	00814	Histolab, Gothenburg, Sweden
POP-7 Polymer for 3730/3730xl DNA Analyzers	4363929	Applied Biosystems, Foster City, CA, USA
RNaseOUT Recombinant Ribonuclease Inhibitor	10777-019	Invitrogen, Carlsbad, CA, USA
Sodium citrate tribasic dehydrate	71402	Sigma-Aldrich, St. Louis, Missouri, USA
Streptavidin-Peroxidase	50-209Z	Invitrogen, Carlsbad, CA, USA
Trizma base	T1503	Sigma-Aldrich, St. Louis, Missouri, USA
UltraPur DNase/RNase-Free Distilled Water	10977-023	Invitrogen, Carlsbad, CA, USA
φX 174-Hae III digest	3405A	Takara Biotechnology Shiga, Japan

## 2.3 Commercial kits

3,3'-Diaminobenzidine (DAB) tetrahydrochloride-plus kit substrate for horseradish peroxidase, cat.no: 0020020 Invitrogen, Paisley, UK

- Substrate buffer concentrate (20X)
- DAB chromogen concentrate (20X)
- Concentrate hydrogen peroxide (20X)
- Concentrate DAB enhancer (20X)

BigDye® Direct Cycle Sequencing Kit, cat.no: 4458687, Applied Biosystems, Foster City, CA, USA

- BigDye® Direct PCR Master Mix
- Sequencing Master Mix
- M13 Fwd Primer
- M13 Rev Primer
- Control DNA CEPH 1347-02

BigDye XTerminator® Purification Kit, cat.no: 4376484, Applied Biosystems, Foster City, CA, USA

- XTerminator™ Solution
- SAM™ Solution

Mouse on Mouse (M.O.M.) Basic Kit, cat.no: MBK-2202, Vector Laboratories, Burlingame, CA, United States

- M.O.M. Protein Concentrate
- Mouse Ig Blocking Reagent
- M.O.M.™ Biotinylated Anti-Mouse IgG Reagent

SuperScript® III Reverse Transcriptase, cat.no: 18080-044, Invitrogen, Carlsbad, CA, USA

- SuperScript III Reverse Transcriptase 200U/μL
- 5X first-strand buffer
- 100 mM Dithiothreitol (DTT)

Qubit dsDNA HS Assay Kit, cat.no: Q32854, Invitrogen, Carlsbad, CA, USA

- Qubit dsDNA HS Reagent
- Qubit dsDNA HS Buffer
- Qubit dsDNA HS Standard #1
- Qubit dsDNA HS Standard #2

Quick-RNA MiniPrek Kit, cat.no: R1054S, Irvine, CA, USA

- Zymo-Spin™ IICG Columns
- DNase/RNase-Free Water
- RNA Lysis Buffer
- DNase I Set

- RNA Prep Buffer
- Spin-Away™ Filter
- RNA Wash Buffer
- Collection Tubes

## 2.4 Solutions prepared in the lab

### Agarose gel (1,5% agarose,) Sub-Cell Model 192

- 350ml 1x TAE buffer
- 5,25g Certified PCR Agarose, cat.no: 161-3105, Bio-Rad Laboratories, Hemel Hempstead, Herts, United Kingdom
- 35,0µl GelRed Nucleic Acid Gel Stain, cat.no: 41003-1, Biotium, Hayward, CA

Mix the agarose powder and TAE buffer and heat the mixture to boiling point in a microwave oven. Cool the solution in room temperature to approximately 65°C before adding GelRed. Pour the solution in a gel chamber and let it polymerize for about 30min. Store the gel in 4°C wrapped in plastic foil if not immediately used.

### Agarose gel (1,5% agarose,) Wide Mini-sub cell GT

- 200ml 1x TAE buffer
- 3,0g Certified PCR Agarose, cat.no: 161-3105, Bio-Rad Laboratories, Hemel Hempstead, Herts, United Kingdom
- 20,0µl GelRed Nucleic Acid Gel Stain, cat.no: 41003-1, Biotium, Hayward, CA

The gele is prepared as described above

### BigDye XTerminator Purification Solution, 1 well

- 10µl XTerminator™ Solution, cat.no: 4376484, Applied Biosystems, Foster City, CA, USA
- 45µl SAM™ Solution, cat.no: 4376484, Applied Biosystems, Foster City, CA, USA

### 0,1% Bromophenol blue gel loading buffer, 25ml

- 0,025g Bromophenol blue, cat.no: 161-0404, Bio-Rad Laboratories, Hemel Hempstead, Herts, United Kingdom
- 5,0g Ficoll PM 400, cat.no: F4375, Sigma-Aldrich, St. Louis, Missouri, USA
- 25ml 1x TAE buffer

Mix the components and vortex the tube about 24 hours to obtain a homogenous solution

### 0.1% BSA

- 10ml dH<sub>2</sub>O
- 1g Bovine Serum Albumin, A9418-10G, Sigma-Aldrich, St. Louis, Missouri, USA

Add 1g of Bovine serum Albumin (BSA) in 10ml dH<sub>2</sub>O

10mM Citrate buffer, 1l

- 2,94g Sodium citrate tribasic dehydrate, cat.no: 71402, Sigma-Aldrich, St. Louis, Missouri, USA
- 1000ml dH<sub>2</sub>O

Mix the solution until the sodium is dissolved. Adjust the pH to 6.0 with 1N HCl measuring the pH with S400 SevenExcellence. Store the solution at room temperature for 3 months or at 4°C for longer storage.

DAB substrate

- 1ml dH<sub>2</sub>O
- 1 drop Substrate buffer concentrate (20X), cat.no: 0020020 Invitrogen, Paisley, UK
- 1 drop DAB chromogen concentrate (20X), cat.no: 0020020 Invitrogen, Paisley, UK
- 1 drop Concentrate hydrogen peroxide (20X), cat.no: 0020020 Invitrogen, Paisley, UK

Add 1 drop of Substrate buffer concentrate, 1 drop DAB chromogen concentrate and 1 drop concentrate hydrogen peroxide in 1ml dH<sub>2</sub>O

10uM dNTP mix, 100µl

- 10µl dATP, cat.no: 4026, Takara, Shiga, Japan
- 10µl dGTP, cat.no: 4027, Takara, Shiga, Japan
- 10µl dCTP, cat.no: 4028, Takara, Shiga, Japan
- 10µl dTTP, cat.no: 4029, Takara, Shiga, Japan
- dH<sub>2</sub>O up to 100µl.

Mix the components and store at -20°C.

70% Etanol, 1l

- 700ml Etanol Absolutt prima, Kemetyl Norge AS, Vestby, Norway
- dH<sub>2</sub>O up to 1000ml

M.O.M Biotinylated Anti-Mouse IgG Reagent

- 2,5ml M.O.M Diluent, Mouse on Mouse (M.O.M.) Basic Kit, cat.no: MBK-2202, Vector Laboratories, Burlingame, CA, United States

- 10µl Biotinylated Anti-Mouse IgG Reagent, Mouse on Mouse (M.O.M.) Basic Kit, cat.no: MBK-2202, Vector Laboratories, Burlingame, CA, United States

Add 10µl stock solution in 2,5ml M.O.M diluent. How to make M.O.M diluent is explained below

M.O.M Diluent

- 1ml PBS
- 80µl M.O.M Protein Concentrate, Mouse on Mouse (M.O.M.) Basic Kit, cat.no: MBK-2202, Vector Laboratories, Burlingame, CA, United States

Add 80µl Protein concentrate in 1ml PBS

M.O.M Mouse Ig Blocking reagent

- 2,5ml PBS
- 2 drops Mouse Ig Blocking Reagent, Mouse on Mouse (M.O.M.) Basic Kit, cat.no: MBK-2202, Vector Laboratories, Burlingame, CA, United States

Add 2 drops of stock solution in 2,5ml PBS

Peroxidase Blocking solution, 41ml

- 40ml Metanol, cat.no: 106009, Merck KGaA, Darmstadt, Germany
- 1ml Hydrogen peroxide solution, cat.no: 216763, Sigma-Aldrich, St. Louis, Missouri, USA

1x Sequencing buffer, 1l

- 100ml 3730 Running Buffer (10X), cat.no: 4335613, Applied Biosystems, Foster City, CA, USA
- dH<sub>2</sub>O up to 1000ml

Mix the components and store the buffer at 4°C.

50x TAE buffer, 1l

- 242g Trizma base, cat.no: T1503, Sigma-Aldrich, St. Louis, Missouri, USA
- 100 ml 0,5 M EDTA (pH = 8,0) EDTA disodium salt, cat.no:100935V, BDH, VWR International, Radnor, PA, USA
- 57,1ml Acetic acid (glacial), cat.no: 100063, Merck KGaA, Darmstadt, Germany
- dH<sub>2</sub>O up to 1000ml

## 2.5 Antibodies

Antigen	Catalog number	Supplier/Manufacturer	Antibody type
Biotinylated Goat Anti-Rabbit	BA-1000	Vector Laboratories, Burlingame, CA, USA	Conjugated Biotinylated Secondary Antibody
Epidermal growth factor receptor	ab2430	Abcam, Cambridge, UK	Rabbit polyclonal
Estrogen Receptor $\alpha$	06-935	Millipore Corporation, Billerica, MA, USA	Rabbit, Polyclonal
Human epidermal growth factor receptor 2	06-562	Millipore Corporation, Billerica, MA, USA	Rabbit, Polyclonal
Keratin 18	61028	Progen Biotechnik, Heidelberg, Germany	Mouse, Monoclonal
Keratin 5	PRB-160P	Covance, New Jersey, USA	Rabbit, Polyclonal
Keratin 6	PRB-169P	Covance, New Jersey, USA	Rabbit, Polyclonal
Ki-67	NCL-Ki67p	Leica Biosystems, Newcastle, UK	Rabbit, Polyclonal
Phospho-p44/42 MAPK (Erk1/2)	4370P	Cell Signaling Technology, Danvers, MA, United States	Rabbit, Monoclonal
Progesterone Receptor	ab131486	Abcam, Cambridge, UK	Rabbit, Polyclonal

## 2.6 Sequencing primers

The primers used for Sanger sequencing were designed using with various online tools and orderer from Eurofins MWG Operon as unmodified DNA Oligos. The oligonucleotides are listed in table 5. In addition to the listed oligonucleotides, the primers had either a M13 forward of M13 revers sequence attached at the 5' end. The primers were attachment with M13 FWD or REV from BigDye Direct Cycle Sequencing Kit.

## 2.7 Various equipment

Equipment	Catalog number	Supplier/Manufacturer
15ml Centrifuge tube	62.554.502	Sarstedt, Nümbrecht, Germany
3730 DNA Analyzer	-	Applied Biosystems, Foster City, CA, USA
50 ml Centrifuge tube	62.547.254	Sarstedt, Nümbrecht, Germany
50 ml Omnifix (60 ml syringe)	461 6520 F	B. Braun Melsungen, Melsungen, Germany
96-Well Non-Skirted Plates	AB-0600	Thermo Scientific, Waltham, MA, USA
Axygen 0.2mL Polypropylene PCR Tube Strips and Domed Cap	PCR-0208-CP-C	Corning ,Tewksbury, MA, USA
BX51 Microscope	-	Olympus America, Melville, New York, USA
Centrifuge MiniSpin	5452 000.018	Eppendorf, Hamburg, Germany
Corning Costar reagent reservoirs	CLS4870	Sigma-Aldrich, St. Louis, Missouri, USA
Cover glass	631-1574	VWR International, Radnor, PA, USA
Curwood Parafilm Laboratory Wrapping Film 4 in. W x 250 ft. L	13-374-12	Thermo Scientific, Waltham, MA, USA
Disposable Scalpels Sterile	0506	Swann-Morton, Sheffield, England
Disposable serological pipets	13-676-10k	Thermo Scientific, Waltham, MA, USA
Domed PCR Cap Strips	AB-0602	Thermo Scientific, Waltham, MA, USA
Eppendorf Centrifuge 5804	5804 000.013	Eppendorf, Hamburg, Germany

Eppendorf Centrifuge 5810R	5811 000.010	Eppendorf, Hamburg, Germany
Eppendorf PCR Clean Safe-Lock Tubes 1,5 ml	0030 123.328	Eppendorf, Hamburg, Germany
GeneGenius Bio Imaging System	-	Syngene, Cambridge, UK
High-Density Polyethylene Staining Jar,	70319	Electron Microscopy Sciences, Hatfield, PA, USA
Illumina High-Speed Microplate Shaker 945195	-	Illumina, San Diego, CA, USA
Leica DM1000 Light Microscope	-	Leica Microsystems, Wetzlar, Germany
MicroAmp Optical 96-Well Reaction Plate with Barcode	4306737	Applied Biosystems, Foster City, CA, USA
Microscopy Slide Box	631.0736	VWR International, Radnor, PA, USA
Moisture Chamber	-	Homemade
NanoDrop 1000 Spectrophotometer	-	Thermo Scientific, Wilmington, DE, USA
PIPETBOY pro	156 404	INTEGRA Biosciences, Zizers, Switzerland
Plate Septa, 96 well	4315933	Applied Biosystems, Foster City, CA, USA
Powerpac 300	-	Bio-Rad Laboratories, Hemel Hempstead, Herts, UK
Precision Wipes Tissue Wipers	75512	Kimberly-Clark Professional, Roswell, GA, USA
Puradisc FP 30 Syringe Filter, cellulose acetate, 0.45 µm, sterile	10462100	GE Healthcare Bio-Sciences, Uppsala, Sweden
Qubit 2.0 Fluorometer	Q32866	Invitrogen, Carlsbad, CA, USA
Qubit Assay Tubes	Q32856	Invitrogen, Carlsbad, CA, USA
Rice cooker	-	Logik
S400 SevenExcellence™ pH/mV	S400	Mettler Toledo, Greifensee, Switzerland
Serological pipettes	86.1251.001	Sarstedt, Nümbrecht, Germany
Spectrafuge Mini Centrifuge	C1301	Labnet, Edison, NJ, USA
Staining Dish, Green	62541-01	Electron Microscopy Sciences, Hatfield, PA, USA
Staining insets	1205/4	Glaswarenfabrik Karl Hecht GmbH & Co KG - "Assistent", Rhön, GERMANY
Sterile Pasteur pipette	135138	LP italiana spa, Milano, Italy
Sub-Cell Model 192	-	Bio-Rad Laboratories, Hemel Hempstead, Herts, UK
Thermaks B8023	-	Thermaks, Bergen, Norway
Tissue-Tek Slide Staining Set	62540-01	Electron Microscopy Sciences, Hatfield, PA, USA
Veriti 96-Well Thermal Cycler	4375786	Applied Biosystems, Foster City, CA, USA
Vortex Mixer SA8	-	Bibby Scientific, Staffordshire, UK
Wide Mini-sub cell GT	-	Bio-Rad Laboratories, Hemel Hempstead, Herts, UK

## 2.8 Database and software

Catalogue Of Somatic Mutations In Cancer (COSMIC), Wellcome Trust Sanger Institute, Cambridgeshire, UK

Cell<sup>^</sup>F imaging systems, Olympus America, Melville, New York, USA GeneSnap, Syngene, Cambridge, UK

Excel 2007, Microsoft, Redmond, Washington, USA

MAFFT, Multiple alignment program for amino acid or nucleotide sequences v.7 (<http://mafft.cbrc.jp/alignment/server/>)

National Center for Biotechnology Information (NCBI), Bethesda, MD, USA

Primer3Plus, Whitehead Institute for Biomedical Research, Cambridge, MA, USA

Primer-BLAST, National Center for Biotechnology Information, Bethesda, MD, USA

R package, v.3.0.2, Department of Statistics of The University of Auckland, New Zealand

Sequencing Analysis Software v5.4, cat.no: 4360967, Applied Biosystems, Foster City, CA, USA

NanoDrop Software v.3.8.1, Thermo Scientific, Waltham, MA, USA

## 3.0 Methods

This chapter describes laboratory procedures used in this master thesis. Methods executed by other members of the research group are also briefly described when relevant to this thesis. If not otherwise is written, I have conducted the experiment.

### 3.1 Mouse model

The impact of increased GLI1 expression on breast cancer formation was addressed by expressing GLI1 in mammary gland epithelial tissue in transgenic mice. The GLI1 expression was induced at the age of 3 weeks in mice with mixed genetic background, of the C57BL/6 and FVB/N strains. The mammary gland tumors developed with long latency. The mice had up to four transgenic alleles, MMTVrtTA, TREGLI1, Lgr5-EGFP-CreERT2 and p53<sup>+/-</sup> (described in section 1.2.1) The mice were genotyped by PCR. The Lgr5-EGFP-CreERT2 alleles was included to enable studies of Lrg5 expression pattern in GLI1 induced mammary gland tumors, but is not of relevance for this thesis. The heterozygote p53 allele was introduced to mimic the p53 status in human breast cancer patients.

Orthotopic serial transplantations of four GLI1 induced tumors were performed for more than 10 generations in NOD/SCID mice to address tumor progression and secure a continued supply of fresh tumor material. This first transplantation was called 0<sup>th</sup> generation. When the tumor had grown to 0.8-1.0cm<sup>3</sup> size, a piece (1-2mm<sup>3</sup>) of the tumor was inserted in a new NOD/SCIS mouse, called 1<sup>th</sup> generation. This continued for more than 10 generations and the 10<sup>th</sup> generation was called “late tumor” in this thesis. Four series were established named A, B, C and D, respectively. The mouse model and the serially transplantations were established at Karolinska Institutet, Stockholm, Sweden.

### 3.2 Exome sequencing

Exome sequencing of GLI1 induced mammary tumors was performed at Science for Life (SciLife) in Uppsala, Sweden, as a collaborative project with Prof. Toftgårs's group at Karolinska Institutet, Stockholm, Sweden. The exome sequencing was done using SureSelect Mouse All Exon Kit from Agilent Technologies, Santa Cruz, CA, USA with SOLiD sequencing technology. The short DNA fragments obtained from the exome sequencing were assembled by bioinformaticians at SciLife using the mm9 NCBI Build 37 C57BL/6 mouse

genome from UCSC Genome Bioinformatics, Santa Cruz, CA, USA as reference. Exome sequencing was performed on four primary GLI1 induced mammary tumors and the corresponding serial transplanted late tumors. Liver DNA from the mouse where the primary tumor arose was also sequenced, to be used as reference for the identification of somatic mutations in the GLI1 induced mammary gland. The exome sequencing data were processed in two different ways, described in chapter 3.8.

### 3.3 Immunohistochemistry

Immunohistochemistry (IHC) was performed on primary and late mammary gland tumor tissue to study the expression pattern of specific proteins. In cases where there were differences between primary and late tumor staining patterns, two intermediate generations (0<sup>th</sup> and 5<sup>th</sup> generation) were also analyzed by IHC.

#### 3.3.1 Antibody selection

IHC staining was performed with antibodies against cytokeratin 5 (K5), K6, K18, Ki-67, human epidermal growth factor receptor 2 (Her2), phosphorylated extracellular signal-regulated protein kinases 1 and 2 (p-Erk1/2) epidermal growth factor receptor (Egfr), estrogen receptor alpha (ER-alpha) and progesterone receptor (PR).

The H&E, K5, K6 and parts of K18 and Ki-67 IHC staining were not performed by me, but has previously been stained by other members of the group.

#### 3.3.2 Fixation, paraffin embedding and sectioning

The tumor tissue was surgically harvested from the mouse mammary gland. The tissue was fixed in 4% paraformaldehyde or FEKETE's fixative to preserve the tissue<sup>40</sup>. The tumor tissue has been harvested in Stockholm, fixed and sectioned by other members of the group. I got the sectioned tissue slides for IHC analyses.

#### 3.3.3 Immunohistochemistry protocol

Two different protocols were used with IHC selected according to the origin of the primary antibody, either with rabbit or with mouse antibody. The protocols were performed according to Vector laboratories<sup>148</sup> and described by Fiaschi et al. 2007<sup>40</sup>.

The standard protocol can be used with primary antibodies of any origin, except mouse. The standard protocol were used to stain with antibodies against ER- $\alpha$ , Her2, Egfr, p-Erk1/2, PR, K5, K6 and Ki-67. The other protocol are called Mouse-on-Mouse (M.O.M) since the primary antibody is from mouse and used on mouse tissue. M.O.M was used to stain with K18 antibody.

The protocol can be divided into 8 steps: deparaffinization, blocking, primary antibody incubation, biotinylated secondary antibody incubation, conjugate incubation, detection of stained celles, nuclear staining and mounting.

The section slides were defaraffinized by incubation for 1 hour in Thermaks heating cabinet. Rehydration of the sections were done in 15 minutes in Xylen twice, through graded alcohol series, twice in 99%, 95% and 70% ethanol for 5 minutes and rinsed in distilled water for 5 minutes in Tissue-Tek Slide Staining. 10mM citrate buffer pH6.0 where used for antigen unmasking. The slides were cooked for 40 minutes in staining jar with citrate buffer with the use of a rice cooker and cooled off for 10 minutes before washing in PBS for 2 minutes 3 times to finish off the deparaffinizing step.

Incubation for 30 minutes with Peroxidase Blocking solution with methanol and hydrogen peroxide were used for blocking endogenous peroxidase activity.

Endogenous peroxidase was blocked with 2,5% hydrogen peroxide in methanol for 30 minutes. The sections were washed 3 times in the standard protocol and 2 times with M.O.M protocol for 2 minutes in PBS (washing step). As a second blocking step 1,5% goat serum was diluted in PBS and 100 $\mu$ l applied to each section separated by wiping the slides with precision wipes. The incubation was done in a moisture chamber to prevent dehydration of the slides. However, for M.O.M protocol the Mouse on Mouse (M.O.M.) Basic Kit was used for the second blocking step due to difficulty distinguishing between mouse primary antibody and endogenous mouse immunoglobins in the tissue. The M.O.M kit is designed so secondary antibody only binds to primary antibody and not to the tissue<sup>148</sup>. 100 $\mu$ l M.O.M Mouse Ig Blocking Reagent was added to each section and incubated for 1 hour to reduce unspecific binding of secondary antibody to endogenous tissue immunoglobulins. The sections were washed 2 times for 2 minutes in PBS. Incubation with 100 $\mu$ l M.O.M Diluent was applied to each section and incubated for 5 minutes before blotting off the solution to finish off the blocking step.

Primary antibody dilution was rested to find the concentration with strong enough staining and minimal background. The dilution from the manufacture was first tested, and diluted more or less according to the background and staining intensity. ER-alpha was

diluted 1:1000, Egfr 1:10.000, p-Erk1/2 1:400, PR:1:1000, Her2, 1:10.000, K5 1:2000, K6 1:2000 and Ki-67 1:10.000 in 0,1% BSA. K18 was diluted 1:20 in M.O.M diluents. The slides were incubated 1 hour at room temperature on over night at 4°C. As negative control one of the two sections on the slide were incubated with primary antibody, while the other section were just incubated in M.O.M diluents or 0,1% BSA in the standard protocol.

The washing step was repeated, before secondary antibody incubation. 1,5% goat serum and 0,5% biotinylated anti-rabbit secondary antibody was diluted in PBS. 100µl was applied to each section in the standard protocol and incubated for 30 minutes. With M.O.M protocol 100µl M.O.M Biotinylated Anti-Mouse IgG Reagent were applied to each section and incubated for 10 minutes. The washing step was repeated.

For the conjugate step, 2 drops of streptavidin peroxidase were applied to each section and incubated for 10 minutes. The sections were washed 3 times for 2 minutes in PBS.

DAB substrate was used as detection system and 100µl substrate were applied to each sections. The color development was monitored under Leica DM1000 Light Microscope and the slides were developed for 2-3 minutes. DAB substrate, counterstaining of nucleus, dehydration and mounting. DAB was develop 1min for ER- $\alpha$ , Egfr 5min, p-Erk1/2 5min, PR 2min, Her2 3-4min, K5 1-2min, K6 3min, K18 2-3min and Ki-67 1-2min. Distilled H<sub>2</sub>O was used to stop the color development.

Mayers hematoxylin was filtered using syringe and syringe filter and used as counterstain to stain the nucleus, and incubated for 30 seconds before rinsing the slides in tap water. The sections were washed for 30 seconds in PBS before washing in dH<sub>2</sub>O.

The sections were dehydrated in increasing ethanol concentrate for 1 minutes in 70%, 95%, 99,5% twice and xylen twice. Pertex mounting medium was applied on the glass slide with Pasteur pipette and cover glass put on the slide as the mounting step. The slides were left to vent off the xylen odor before the slides were examined and picture taken with BX51 Microscope and Cell<sup>F</sup> imaging system. A suitable section of the slide, with as little noise as possible, was taken picture of. The only setting done in Cell<sup>F</sup> imaging system was adjusting the white balance of the picture. The pictures were taken at 40x and 400x magnification.

The scoring of percentage positively stained tumors cells was estimated approximately for each of the biomarkers. The positively cells was not counted be scored in a appropriate area of the tumor.

### 3.4 DNA isolation

DNA used for Sanger sequencing were isolated with QIAcube, QIAGEN, Venlo, Netherlands according to manufacturer's protocol. DNA isolation was performed with help from Eldri Undlien Due, Department of Genetics. The DNA concentrations were measured with Qubit as described in chapter 3.6.

### 3.5 RNA isolation

RNA was isolated from mammary glands to be used in cDNA synthesis. RNA was isolated using TRIzol based method. One of the samples (sample 17) didn't give any reasonable results with sequencing. Others in the group had also had problem with this sample, so this sample was isolated with a second method, Quick-RNA MiniPrep Kit. Sample 17 isolated with this method are referred to as sample 17 Quick in the rest of this thesis.

RNA isolation was done according to ZYMO research's instruction manual<sup>149</sup>. Quick-RNA MiniPrep is fast-spin column method. 0,02g frozen tumor tissue were lysed with 300µl RNA Lysis Buffer as Sample lysis/homogenization (step 1) in the protocol. Sample clearing and gDNA removal (step 2) and RNA purification (step 3) was performed according to the protocol, and DNase I treatment listed as optional was conducted. The centrifugation was done with Spectrafuge Mini centrifuge.

RNA concentrations were measured on Thermo Scientific NanoDrop 1000 Spectrophotometer according to Thermo Fisher Scientific Users manuals<sup>150</sup>. Nucleic Acid application module was chosen. NanoDrop was first initialized with 1µl water. The RNA concentration was measured in 1µl of eluted sample solution, DNase/RNase-Free water, was used as blank to zero out contributing factor. 1µl of sample was measured and the concentrations were used in cDNA synthesis.

### 3.6 cDNA synthesis

To investigate the expression of the chosen mutations in the mammary gland tumors, Sanger sequencing was performed on complementary DNA (cDNA).

cDNA synthesis was performed with RNA isolated in TRIzol and Quick-RNA MiniPrep Kit. 0,6µg RNA was used as input. Sample 17 Quick had lower than the rest of the samples and input was only 0,27µg compared to the rest of the samples with 0,6µg input.

The enzyme SuperScript® III Reverse Transcriptase was chosen because of its increased specificity, reduced RNase activity and high cDNA yield<sup>151</sup>. The first strand synthesis was performed according to Invitrogen's SuperScript® III Reverse Transcriptase protocol<sup>152</sup>. 0,6 µg RNA, Oligo(dt)<sub>20</sub> and 10mM dNTP mix were used in the first incubation step. A master mix with 5X First-Strand buffer, 0,1M Dithiothreitol (DDT), RNase OUT and 200U/µL SuperScript III Reverse Transcriptase enzyme was used in the second incubation step. Veriti 96-Well Thermal Cycler was used for the incubation steps.

NanoDrop cannot distinguish between RNA, DNA, free nucleotides and other contaminants thus Qubit 2.0 Fluorometer was used to determine the cDNA concentration<sup>153</sup>. The measurements with Qubit 2.0 was performed according to the experimental protocol by Molecular Probes<sup>145</sup> with 1µl cDNA. Qubit dsDNA HS Assay Kit and Qubit Assay Tubes was used with this protocol. The cDNA synthesis was done in multiple runs as it was necessary for more cDNA throughout the experimental work.

### 3.7 Sanger sequencing

To validate the results from whole exome sequencing, Sanger sequencing of cDNA was conducted. cDNA sequencing was chosen because it is also interesting to determine whether or not the mutated genes actually are expressed in mammary gland tissue. In addition, Sanger sequencing was conducted on genomic DNA to validate the Kras mutation.

#### 3.7.1 Selected genes and primer design

22 mutations identified by whole exome sequencing were chosen to be sequenced with Sanger sequencing. First, four mutations identified by exome sequencing in genes commonly associated with breast cancer were chosen to be validated by Sanger sequencing, including Brca2, Brcc3, Kras and Ptch1. Secondly, mutations identified in both primary and late tumors

were chosen, including *Atp13a3*, *Btnl6*, *Commd7*, *Fam179b*, *Lpin1*, *Myom1*, *Sema6d*, *Sfi1* and *Stat6*. Finally, mutations found in the COSMIC Cancer Gene Census database were chosen to be sequenced, including *Cdk4*, *Foxp1*, *Kdm6a*, *Il7r*, *Maml2*, *Psip1* and *Tsc1*. *Ptch1* and *Kras* are also present in COSMIC Cancer Gene Census. An overview of the mutations is listed in table 6 with additional information about the mutation. All the mutations cDNA sequenced are nonsynonymous mutations. Which series the mutations is identified in by exome sequencing is shown in appendix A.

One normal breast tissue sample from mouse was sequenced with each of the primer pairs, to compare the tumor samples against and confirm potential somatic mutations. Mutations found in late tumors and not in primary tumors and genes found in COSMIC Cancer Gene Census database was in addition sequenced with normal breast tissue from a NOD/SCID mouse as control, to exclude mutations that were caused by germline in the NOD/SCID strain.

The primers were designed by using Primer3Plus and Primer-BLAST and ordered from Eurofins MWG Biotech. The primers are listed in table 3 and the table do not show the M13 forward and reverse primer that are attached to each of the 5'-end of the primers. The coding part of the reference sequence was found in National Center for Biotechnology Information (NCBI) and pasted into Primer3Plus. The mutation position identified by analyzing of the exome sequencing data was chosen as sequence target and product size was chosen from 400 to 600 bases. Primer3Plus suggested several primers at different positions in the sequence, and the primers not too close or too far from the mutation base was chosen. The coding part of the reference sequence, forward and reverse primer were pasted in to Primer-BLAST to check the specificity of the chosen primers. Only the parameter "organism" was changed to *mus musculus* in the program. The primers that were eventually chosen were not aligned to other transcripts by Primer-BLAST.

Table 3: The table lists the genes, reference sequences available and position of the coding part of the reference sequences used for Sanger sequencing. The base and amino acid substitutions listed are those found from exome sequencing. Some genes are listed twice, because exome sequencing identified multiple mutated position in that gene. Oligonucleotides used as primers, both forward (FWD) and reverse (REV) direction, are listed for each gene. M13 forward and reverse primer was also attached to each oligonucleotide, but not shown in this list for simplicity. Some genes have multiple primers, indicated by New. Kras gene was sequenced using both cDNA and gDNA. Sequenced Kras with gDNA indicated by NC reference sequence. The G3587T is the genomic mutation in Kras in reference sequence, and G12V is the protein substitution of interest.

Gene Name	Reference sequence	Base substitution	Amino acid substitution	Direction	Oligonucleotide
Atp13a3	NM_001128096 NM_001128094	G1126T	V376L	FWD	ggcaactcacagtaccgtga
				REV	tcattgcagcaggaagagca
Brca2	NM_009765 NM_001081001	A6432T	K2144N	FWD	gcactgcaggtggaactg
				REV	gcactctcactgcttctgt
Brcc3	NM_001166457 NM_145956 NM_001166459	A153T	K51N	FWD	tgcaggcgggtcatcttgag
				REV	gggcttgatggattggaagc
Btl6	NM_030747	G949T	A317S	FWD	gctgatcgtcaggagctaa
				FWD	cacaaacacccccacactct
				FWD New	ggtggagcttgctcaaagga
				REV New	ctctccaggctgcct
Cdk4	NM_009870	G67A	A23T	FWD	tgccactgatatgaaccg
		G151A	V51I	REV	ttcaggagctcgggtaccaga
Commd7	NM_133850	G283A	D95N	FWD	gagacatgcagcagctgaac
	NM_001195390	G70A	D24N	REV	tctctccatctcgtgcagga
Fam179b	NM_177805	A2152T	N718Y	FWD	ccatttggcacatggagcag
				REV	tggctgtccccaaaactgt
Foxp1	NM_001197321	G1399T	A467S	FWD	tgcaaaagacaaaagcgcgc
				REV	gtggcctcgttttggaaac
Il7r	NM_008372	C1253T	P418L	FWD	gtcgtatggcctagtctccc
				REV	ctgagggaagtggagatgggc
Kdm6a	NM_009483	A3467C	N1156T	FWD	taccaggcctcctcattcca
				REV	ttgttcaccagccaatagc
Kras	NM_021284	G35T	G12V	FWD	aactgtgtgtgttggagct
				REV	actgtacacttgctctgact
	NC_000072	G3587T		FWD gDNA	tgggatagctgtcacaagc
				REV gDNA	tcttttcaaagcggctggc
Lpin1	NM_172950 NM_001130412	A1475C	Q492P	FWD	agtggcagccctgtatttcc
	NM_015763	A1574C	Q525P	REV	tccggtattgtggccttcc
Maml2	NM_173776	A250C	S84R	FWD	cccagcagtcctgttgaat
				REV	taagggagatggccgagaca
Myom1	NM_001083934 NM_010867	C647T	T213M	FWD	gtccaccctcaacgagagg
				FWD	gacagcccagactcatggtc
Ovgp1	NM_007696	G1588A	A530T	REV	ctggctgggctggttcttc
				FWD	cctccaatggacagcagtg
Psp1	NM_133948	A496G	T166A	REV	tcaggagaaaaggccagca
				FWD	agcatgtcctcctgtgagc
Ptch1	NM_008957	C888A	N296K	FWD	gacagctgggaggaaatgct
				REV	catcagtaggtagccgctgg
		G265T	V89F	FWD	tggagcagattccaagggg
				REV	cccagaagcagtccaaggt

<b>Sema6d</b>	NM_199239 NM_199241 NM_172537 NM_199238 NM_199240	G482C	S161T	FWD	cggtgctgttccttctggtct
				REV	cgaagggcagatccatctcc
<b>Sfi1</b>	NM_030207	C1066T	R356C	FWD	acctccaaatctgcagggtg
				REV	tcgctgctggaaatgacat
				FWD New	cactggaagcactacatgttgc
				REV New	ctaaagcaagaatacagcaggt
<b>Stat6</b>	NM_009284	G1487A	C496Y	FWD	ttgggcccacaacttct
				REV	gaggaggaaggtccatct
<b>Tsc1</b>	NM_022887	A1118C	Y373S	FWD	tgttgttcagtcacctggg
				REV	gaggaactacaggatccgc

### 3.7.2 PCR amplification

PCR amplification was the first step in Sanger sequencing. BigDye direct cycle sequencing kit protocol <sup>154</sup> was used and the PCR reaction was conducted as described in the protocol. Master mix with BigDye direct PCR master mix from BigDye Direct Cycle Sequencing Kit, designed M13-tailed forward and reverse primer and UltraPur DNase/RNase-Free Distilled Water was applied in MicroAmp® Optical 96-Well Reaction Plate with Barcode and 1 or 2µl (5ng/µl) cDNA was added. Distilled water was adjusted to obtain same volume in each reaction. A negative control was included in every run and processed the same way as the DNA containing samples by replacing the DNA sample volume with distilled water. The plate was sealed with Domed PCR Cap Strips and centrifuged with Eppendorf Centrifuge 5810R. The PCR amplification step was performed as 9700 thermal cycler in the protocol with the use of Veriti 96-Well Thermal Cycler.

BigDye direct cycle sequencing kit protocol <sup>154</sup> recommends that the PCR products should be checked after the PCR reaction on agarose gel to determine the quality and quantity of the PCR product. 2µl PCR product was mixed with 2µl 0,1% Bromophenol blue Gel loading buffer in 96-Well Non-Skirted Plates and sealed with Domed PCR Cap Strips and centrifuged with Eppendorf Centrifuge 5804. Each mixed solution were applied in separate well in 1,5% agarose gel in Sub-Cell® Model 192 or Wide Mini-sub cell GT according to number of samples and the gel size. 2µl φX 174-Hae III digest was used as ladder marker. Powerpac 300 was used to apply electric field in 1X TAE buffer and the gel was run for 30 minutes. GeneGenius Bio Imaging System and GeneSnap software was used to make the band on the gel visible for evaluation.

### 3.7.3 Chain termination sequencing

Cycle sequencing is the chain termination step. Cycle sequencing was performed according to BigDye direct cycle sequencing kit protocol<sup>154</sup>. Sequencing was performed in both forward and reverse direction using Direct Sequencing Master Mix, M13 Fwd Primer and M13 Rev Primer from BigDye Direct Cycle Sequencing Kit. The cycle sequencing reaction was performed as 9700 thermal cycler option in the protocol.

Sequencing products were purified using BigDye XTerminator purification kit according to the manufacturer's protocol<sup>154</sup>. Sterile disposable scalpels were used to cut a little part of the pipette tip to pipette the XTerminator solution. The plate was sealed with Domed PCR Cap Strips. The plate was vortexed for 30 minutes at 2000rpm using Illumina High-Speed Microplate Shaker. The remaining empty wells were filled with deionized water using Corning Costar reagent reservoirs.

### 3.7.4 Capillary electrophoresis and results processing

The capillary electrophoresis was done with 3730 DNA Analyzer. Starting of 3730 DNA Analyzer and Collection Software were done according to Applied Biosystems 3730/3730xl DNA Analyzers User Guide<sup>155</sup>. The buffer reservoir was filled with 1x Sequencing buffer. Septa sealing was used with Plate Septa, 96 well, and assembled as explained in the User Guide. POP-7 Polymer for 3730/3730xl DNA Analyzers was used as the sequencing performance optimized polymer (POP) as recommended<sup>154</sup>.

The Sequencing Analysis software was used for processing the raw data obtained from capillary electrophoresis. The mutated position was inspected manually to determine if the mutation were present. If a mutation was present, the signal intensity of the wild type was compared to the signal intensity of the mutation in the same base position to find the percentage of detected mutation. The overall sequence quality was inspected with considering the signal intensity, signal evenness, and background noise. MAFFT sequence alignment software was used to compare the sequence to the reference sequence and check that the sequenced area is correct.

### 3.8 Statistical analysis/bioinformatics

The assembled sequences from SciLife were processed in two different ways to identify possible point mutations. Mutation calling based on the mm9 NCBI Build37 was named “reference call” in this thesis. The mm9 NCBI Build 37 is pure C57BL/6 genome. The GLI1 induced tumors have mixed genetic background of C57BL/6 and FVB/N. Thus, the mutation calling using only the C57BL/6 reference genome might result in identified mutations due to the mixed genetic background and not due to real mutations. To exclude potentially false positive mutations the assembled sequences were compared to other known mouse genomes and to liver DNA from the same animal as the primary tumor arose in. The results of the liver DNA exon sequencing were processed through the same pipeline as the tumor samples. This point mutation calling method was named “animal specific call” in this thesis. Members of Erik Fredlund's group at Karolinska Institutet, Stockholm, Sweden, performed the filtration and mutation calling.

With the results from the two different mutation calling (animal specific call and reference call), it was of interest to look at how many unique mutations were identified with each of the calling methods and how many unique mutations were common between the two methods. The statistical software package R was used for this purpose. Daniel Nebdal at Department of Genetics helped with the statistical analyses in R.

The results from “reference call” were further used to search for mutated human homologous and if the mutations have been detected in human cancer and human breast cancer. The reference call detected mutated genes were compared to the catalogue of somatic mutations in cancer (COSMIC) database. The COSMIC cancer gene census database indicated which of genes present in reference call data are known to have cause cancer. These analyses were performed by Members of Erik Fredlund's group at Karolinska Institutet.

The MAFFT sequence alignment software was used to do multiple sequence alignment of the sequences obtained from Sanger sequencing and compare these to the reference sequence to check that the correct area had been sequenced. The coding part of the reference sequence, sequences from both forward and reverse direction from Sanger sequencing, from all samples with the same primer pair were compared. The box allowing unusual symbols was ticked off since a position with several bases, like a mutation, are indicated with a different letter than A, C, G or T in the sequencing analysis software. The gap opening penalty was set to the strictest value. No other parameters were altered. The alignment was manually checked to see if it had

correctly identified the reverse directions and reversed them, and that the alignment shows correct sequenced area.

## 4.0 Results

### 4.1 Immunohistochemistry

Immunohistochemistry (IHC) was performed to characterize GLI1 induced mammary tumors, and to study the stability in the serially transplanted tumors. An overview of the result from all IHC and hematoxylin and Eosin (H&E) staining is shown in table 4. Negative controls were negative as anticipated in every sample with the different antibodies.

H&E staining of GLI1 induced mammary tumors from primary and late tumors was performed for tumor characterization (figure 16). The H&E staining was judge by other members of the group and concluded that series A primary tumor was solid carcinoma, series A late, series B primary and late tumors was adenocarcinomas, series C tumors were carcinosroma/anaplastic and series D primary was characterized as carcinosarcoma. Series D late tumor form the 10<sup>th</sup> generation have not been characterized.

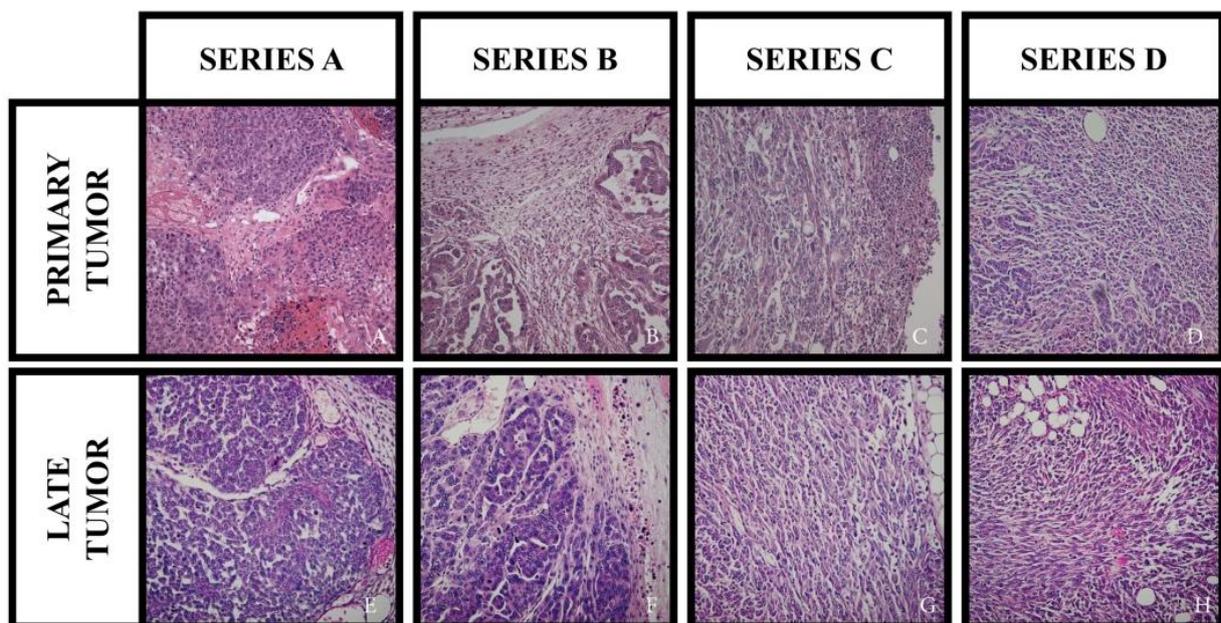


Figure 16: Four series of GLI1 induced mammary tumors stained with hematoxylin and eosin in primary (A-D) and late tumors from the 10<sup>th</sup> generation transplantation (E-H). Magnification 100x (D, F-H) and 200x (A-C, E).

Cytokeratin 5 (K5), K6 and K18 staining were performed to determine if the GLI1 induced mammary gland tumors had luminal or basal-like characteristics. K5 and K6 are basal keratins and K18 is luminal keratin. Pictures showing correct epithelial cell staining with K5, K6 and K18 are shown in appendix B for comparison. Series A was stained positively for K5

and showed a few K6 positively stained cells (figure 17). Series A had very few positive cells for K18 staining but the majority of the tumor cells were negative indicating that the tumors were negative for the K18 marker. Series B, C and D showed staining with the K18 antibody. Series B also showed some weak staining with the K6 antibody in the late tumor. Series C had weak staining for K18 and did not show any staining for K5 and K6 indicating that the tumors were only K18 positive. Series D did not either show any K5 or K6 staining of the tumor cells. The positive K5 staining observed in series D was located to normal cells in a duct. Since there were some differences in K18 staining between primary and late tumor, especially in series A and C, the 0<sup>th</sup> and 5<sup>th</sup> generations from all four series were also stained for K18 (pictures in appendix B). Series A showed no staining either in the 0<sup>th</sup> or 5<sup>th</sup> generation, while series B showed the same number of positive cells throughout the serially transplanted tumors. Some variation in number of positively stained tumor cells with K18 antibody were observed series C, while in series D, the 0<sup>th</sup> generation showed loss of K18 staining which emerged again in the 5<sup>th</sup> generation and in the late tumor. To summarize, Series A had tumor cells stained with K5 and K6 basal antibody, while series B, C and D were stained with the luminal marker K18. In addition series B late tumor was stained with K6.

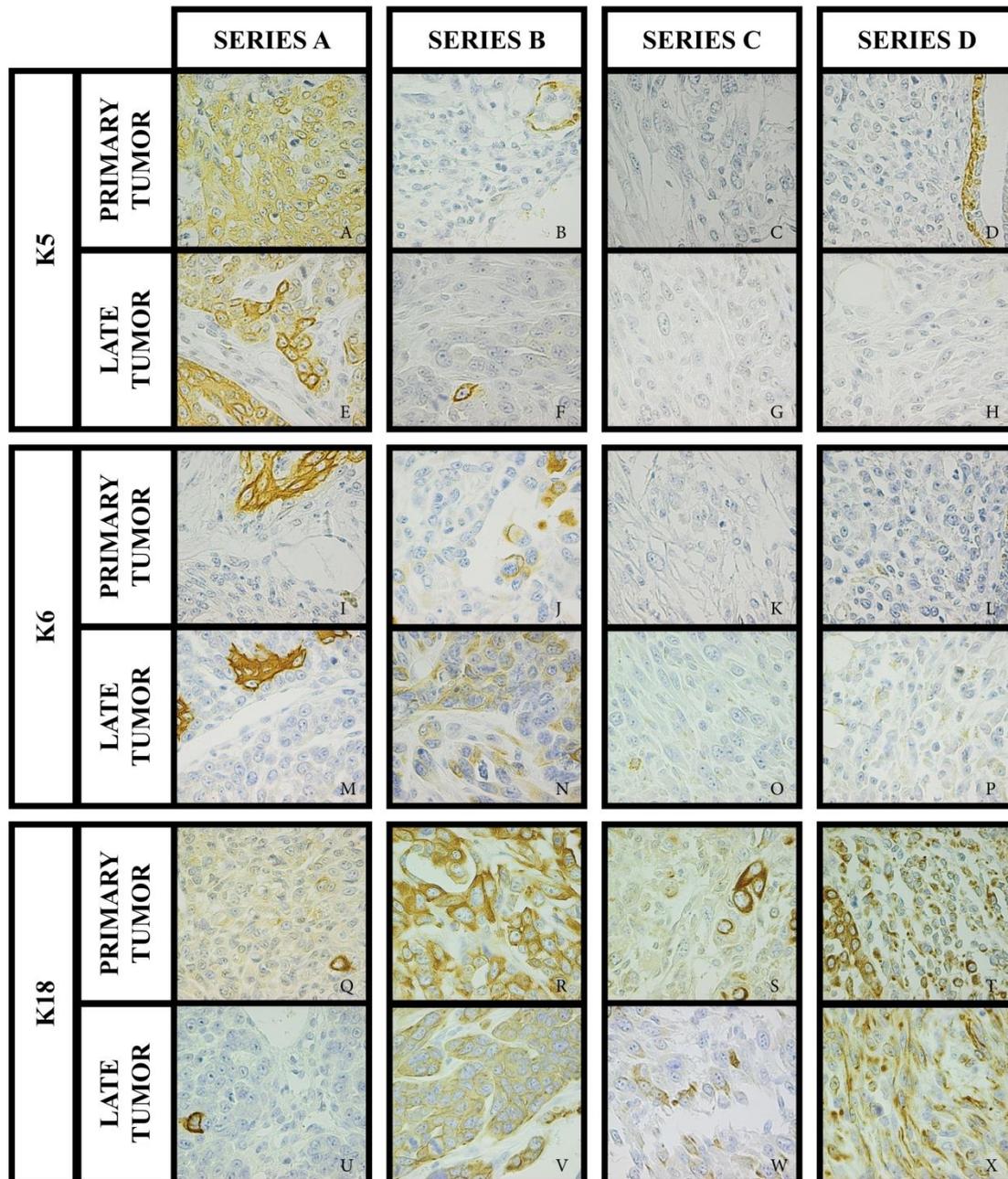


Figure 17: Four series of GLI1 induced mammary gland tumors stained with K5 (A-H), K6 (I-P) and K18 (Q-X) antibodies in primary tumors (A-D, I-L, Q-T) and late tumors from the 10<sup>th</sup> generation transplantation (E-H, M-P, U-X). Magnification 400x.

IHC with the proliferation marker Ki-67 was performed to study the proliferation of the GLI1 induced mammary gland tumors and the stability in serially transplantations. Ki-67 staining showed increasing number of positively stained cells between primary and late tumors from the 10<sup>th</sup> generation transplantation, so 0<sup>th</sup> and 5<sup>th</sup> generation in all four series were in addition stained with Ki-67 antibody (figure 18). The Ki-67 antibody staining showed increasing

number of positively stained cells over the course of the serial transplantations in all four series.

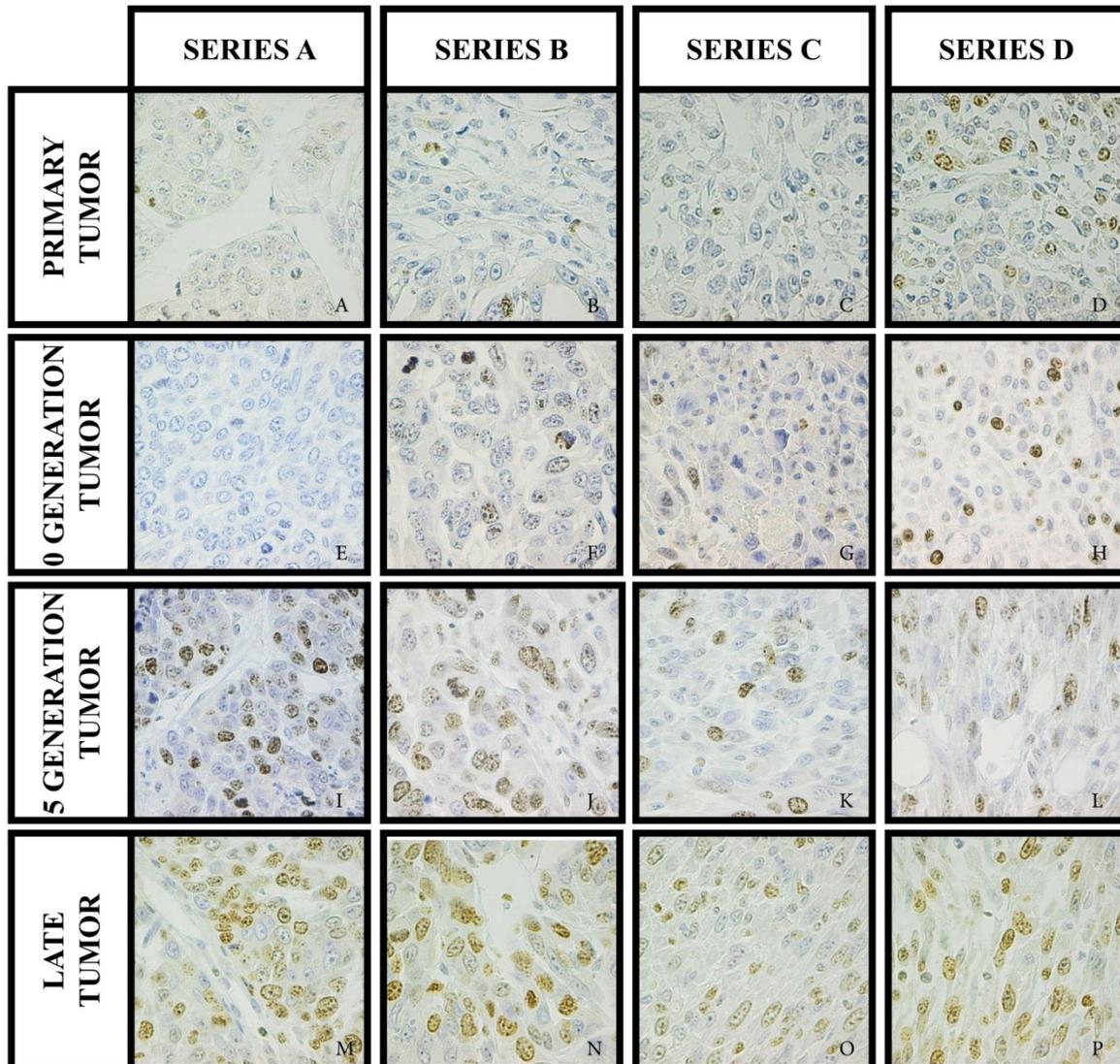


Figure 18: Four series of GLI1 induced mammary tumors stained with Ki-67 (A-P) antibody from primary tumors (A-D), 0<sup>th</sup> (E-H), 5<sup>th</sup> generation (I-L) and late tumors from the 10<sup>th</sup> generation transplantation (M-P). Magnification 400x.

Similarly, IHC using antibodies to receptors Her2, ER-alpha and PR was performed to characterize the GLI1 mammary gland tumors with respect to the hormone receptor status ER-alpha, PR and the growth factor receptor Her2. ER-alpha staining showed a few positive cells in each tumor, except in primary tumor series A (figure 19). The four primary tumors were negative for PR staining. However, the late serially transplanted tumors showed cytoplasmic staining and some few nuclear positive cells. Primary tumor in series D was not

available for PR staining. Her2 was negative stained in all tumors from all four series. Positive staining with Her2 antibody of a normal duct in tumor tissue is shown in appendix B.

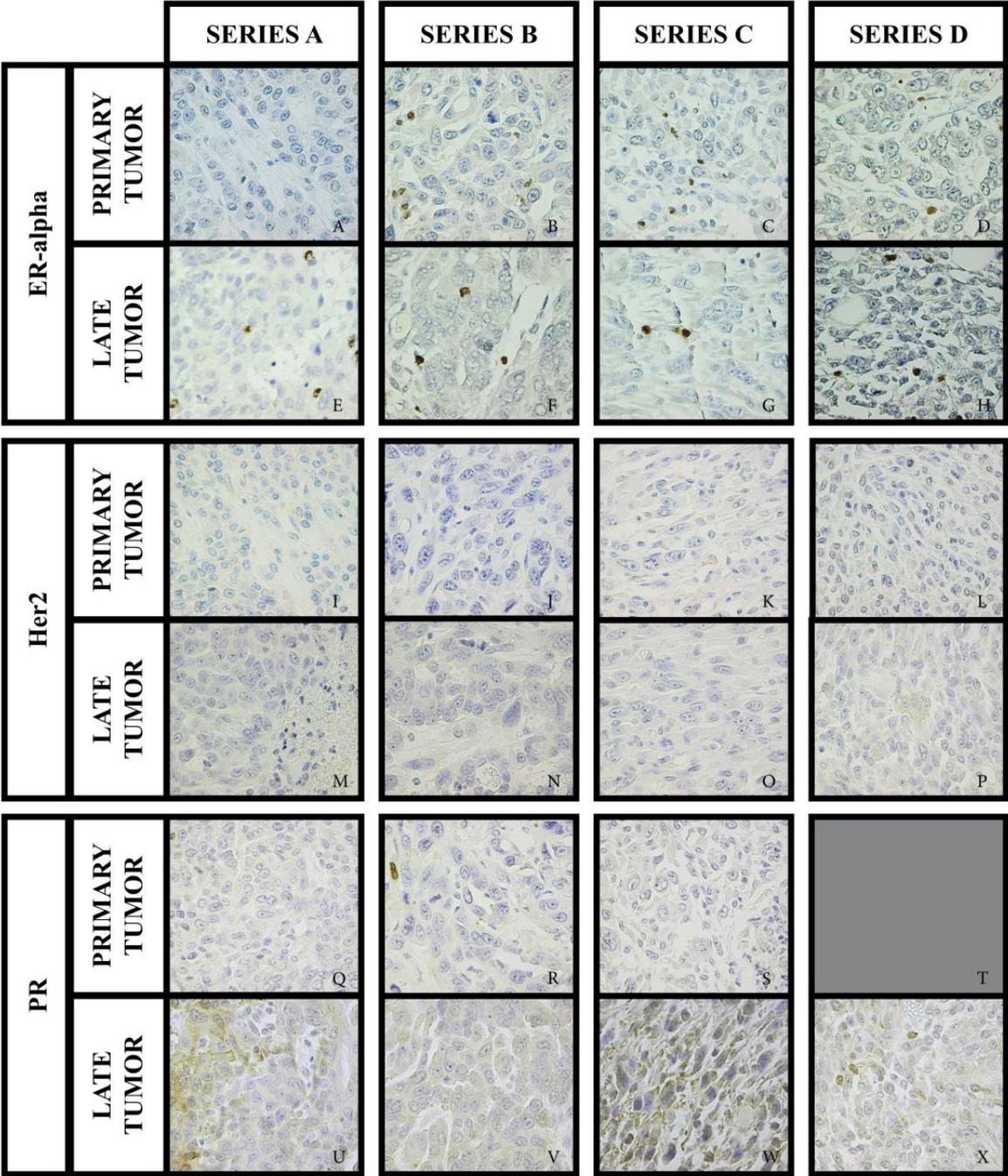


Figure 19: Four series of GLI1 induced mammary tumors stained with ER-alpha (A-H), Her2 (I-P) and PR (Q-X) antibody from primary tumors (A-D, I-L, Q-S) and late tumors from the 10<sup>th</sup> generation transplantation (E-H, M-P, U-X). Primary tumor for series D (panel T) was not available. Magnification 400x.

Staining with antibody against Egfr showed that primary and late tumors did not express detectable levels of Egfr (figure 20). Only a few positive cells were detected at the edges of the tumors. Positive staining of Egfr in a normal duct in a tumor tissue sample is shown in appendix B. Staining for p-Erk1/2 showed more cytoplasmic than nuclear staining. Nuclear staining was found in series C as well as in series B and D late tumors. Tumors in series A and primary tumor in series B were negative for p-Erk1/2 staining. Series D primary tumor showed only cytoplasmic staining. p-Erk1/2 staining showed difference number of positively stained cells between primary and late tumors from the 10<sup>th</sup> generation transplantation, so 0<sup>th</sup> and 5<sup>th</sup> generation in all four series were in addition stained with p-Erk1/2 (appendix B). Staining of tumors from four different generations in series B showed that the amount p-Erk1/2 gradually increased. Series C and D tumors had weak p-Erk1/2 staining in the 0<sup>th</sup> generation and then staining increased throughout the serial transplantations. Some of the tumors showed most p-Erk1/2 staining at the edges of the tumor (series B: 0<sup>th</sup>, 5<sup>th</sup> generation, late, series C: 0<sup>th</sup> generation, series D: primary, late) while others had negative tumors edges even though the tumors were positive for p-Erk1/2 (series C: 5<sup>th</sup> generation, late, series D: 0<sup>th</sup> generation), data not shown.

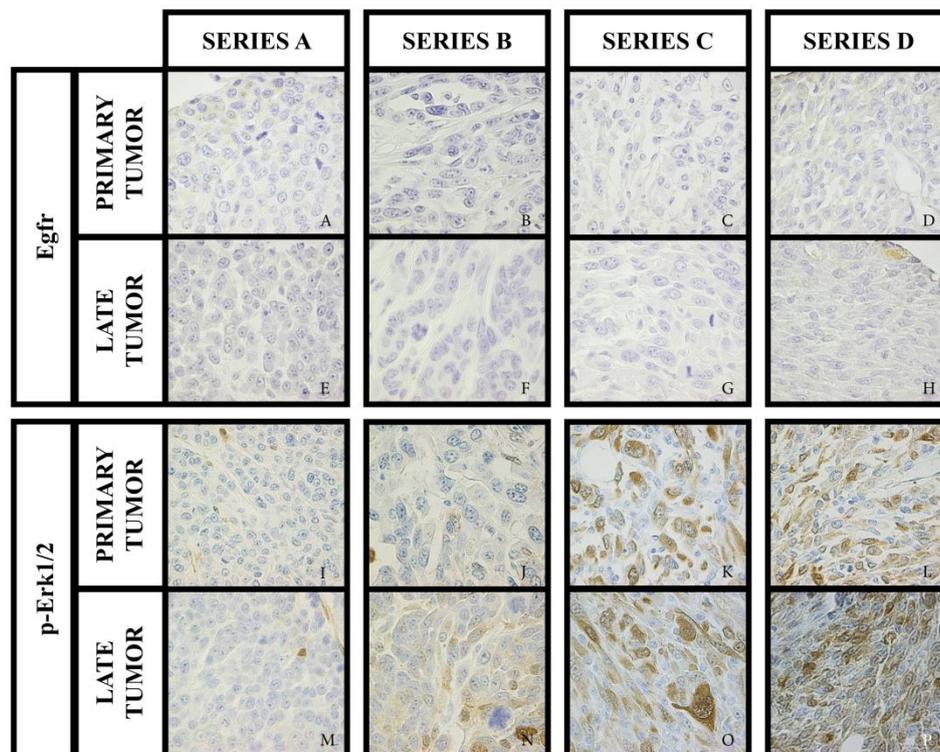


Figure 20: Four series of GLI1 induced mammary tumors stained with Egfr (A-H) and p-Erk1/2 (I-P)) antibody from primary tumors (A-D, I-L) and late tumors from the 10<sup>th</sup> generation transplantation (E-H, M-P). Magnification 400x.

Table 4: An overview of immunohistochemistry results from four series of GLI1 induced primary mammary tumors and late tumors from the 10<sup>th</sup> generation transplantation. Scoring: negative when <10% of the tumor cells have positive staining, 1+ is cells that have 10-25% positive staining, 2+ 25-50%, 3+ 50-75% and 4+ > 75% positively stained cells. The scoring is based on estimated percentage of positively stained tumor cells of each marker in an area of the tumor.

	Series A Primary	Series A Late	Series B Primary	Series B Late	Series C Primary	Series C Late	Series D Primary	Series D Late
H&E	Solid carcinoma	Adeno-carcinoma	Adeno-carcinoma	Adeno-carcinoma	Carcino-srcoma/anaplastic	Carcino-srcoma/anaplastic	Carcino-sarcoma	NA
K5	++++	+++	-	-	-	-	-	-
K6	+	+	-	++	-	-	-	-
K18	-	-	++++	++++	+	+	+++	+++
Ki-67	-	+++	-	+++	-	++	-	++
Her2	-	-	-	-	-	-	-	-
ER-alpha	-	-	-	-	-	-	-	-
PR	-	-	-	-	-	-	-	x
Egfr	-	-	-	-	-	-	-	-
p-Erk1/2	-	-	-	+	+	+	+	+

## 4.2 Exome sequencing

The exome sequencing data were analyzed in two different ways to identify somatic mutations. Exome sequencing results aligned with reference sequences and liver DNA exome sequencing (animal specific call) identified less unique mutations than the exome sequencing results aligned with a reference genome (reference call). These analyses identified a total of 1727 unique mutations using the two calling methods combined. Only 106 of the unique mutations were overlapping (figure 21).

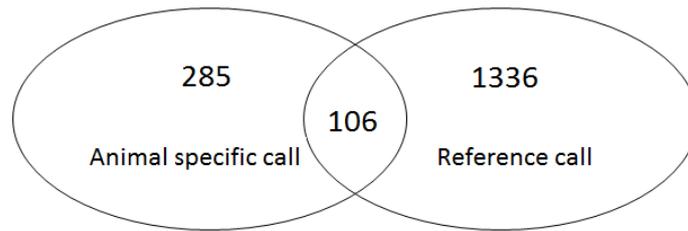


Figure 21: Venn diagram showing the number of unique mutations from using two different mutation calling methods. 391 unique point mutations were found with the animal specific calling and 1442 unique point mutations were found with the reference call. 106 point mutations were identified by both methods.

Table 5 lists the number of somatic mutations found in each of the tumors, and overlapping mutations between the different samples. The diagonal shows the number of mutations in each tumor sample. From the table it is clear that the late tumors have many more mutations than in the primary tumors.

Table 5: Number of overlapping mutations identified by animal specific call between series A, B, C and D, from both primary and late tumors from the 10<sup>th</sup> generation transplantation.

	Series A Primary	Series A Late	Series B Primary	Series B Late	Series C Primary	Series C Late	Series D Primary	Series D Late
Series A Primary	<b>10</b>	1	0	0	0	0	0	0
Series A Late	1	<b>126</b>	0	23	0	29	0	13
Series B Primary	0	0	<b>18</b>	7	0	0	0	0
Series B Late	0	23	7	<b>56</b>	0	16	0	7
Series C Primary	0	0	0	0	<b>11</b>	3	1	1
Series C Late	0	29	0	16	3	<b>193</b>	1	20
Series D Primary	0	0	0	0	1	1	<b>9</b>	4
Series D Late	0	13	0	7	1	20	4	<b>65</b>

### 4.3 Sanger sequencing

Of 31 mutations in the GLI1 induced tumors identified by exome sequencing only 58% of the mutations were identified by cDNA sequencing. The results from Sanger sequencing is listed in table 6 along with the results from exome sequencing. In which of the series the mutations are found in, are listed in appendix A. *Btln6* expression was not in any of the samples. All

nine genes present in COSMIC Cancer Gene Census database were controlled-sequenced using normal NOD/SCID mammary tissue with both gDNA and cDNA sequencing. The control sequencing with NOD/SCID samples were negative for the mutation. An overview over which genes this applies to, is shown in appendix A.

Table 6: The table lists the percentage of the mutations that have been detected with genomic exome sequencing and cDNA sequencing from GLI1 induced mammary tumors tissue and normal liver tissue. The gray area represents unavailable data. Reference and animal specific calling of mutation have negatively or positively identified the mutation from exome sequencing data.

Gene	Sample	Exome	Animal Specific call	Reference call	cDNA
Atp13a3	Primary	47	+	+	67
	Late	40	+	+	50
	Liver	0			
Brca2	Primary	0	-	-	0
	Late	38	-	+	0
	Liver	0			
Brcc3	Primary	0	-	-	0
	Late	15	-	+	0
	Liver	0			
Btl6	Primary	26	+	+	Not expressed
	Late	50	+	+	Not expressed
	Liver	0			
Cdk4 G151A	Primary	2	-	-	0
	Late	8	+	-	0
	Liver	2			
Cdk4 G67A	Primary	2	-	-	0
	Late	12	+	-	0
	Liver	1			
Comm7	Primary	17	+	-	33
	Late	27	+	+	35
	Liver	0			
Fam179b	Primary	58	+	+	86
	Late	79	+	+	100
	Liver	0			
Foxp1	Primary	0	-	-	0
	Late	11	+	-	?
	Liver	0			
Il7r	Primary	0	-	-	0
	Late	13	+	-	0
	Liver	0			
Kdm6a	Primary	15	+	-	11
	Late	0	-	-	0
	Liver	0			
Kras	Primary, C	9	+	-	63
	Late, C	30	+	+	75
	Liver, C	0			
	Primary, D	46	+	+	78
	Late, D	78	+	+	88

	Liver, D	1			
Lpin1	Primary	53	+	-	53
	Late	80	+	+	90
	Liver	0			
Maml2	Primary	0	-	-	0
	Late	16	+	-	8
	Liver	0			
Myom1	Primary	5	+	-	0
	Late	10	+	-	13
	Liver	0			
Psip1	Primary	10	-	-	0
	Late	21	+	-	0
	Liver	2			
Ptch1 G265T	Primary	23	-	-	0
	Late	26	+	-	0
	Liver	17			
Ptch1 C888A	Primary	18	-	+	0
	Late	1	-	-	0
	Liver	0			
Sema6d	Primary	6	+	-	0
	Late	16	+	-	0
	Liver	2			
Sfi1	Primary	3	+	-	0
	Late	6	+	-	0
	Liver	2			
Stat6	Primary	7	+	-	21
	Late	31	+	+	42
	Liver	0			
Tsc1	Primary	0	-	-	0
	Late	9	+	-	14
	Liver	0			

18 of the 31 mutations selected to be validated by Sanger sequencing from both the reference call and animal specific call data were confirmed with cDNA; 13 mutations were not validated, giving a total of 58% verified mutations (figure 22). Mutations identified by both the reference and animal specific call were validated in 100% of the cases. Sanger sequencing validated 18 of the 28 mutations identified by using the animal specific call method; hence 10 of 28 mutations were not validated, resulting in 64% validation of this method by Sanger sequencing. Finally, 9 of 12 mutations which were identified by the reference call method were confirmed whereas 3 were disclaimed, giving 75% validation rate.

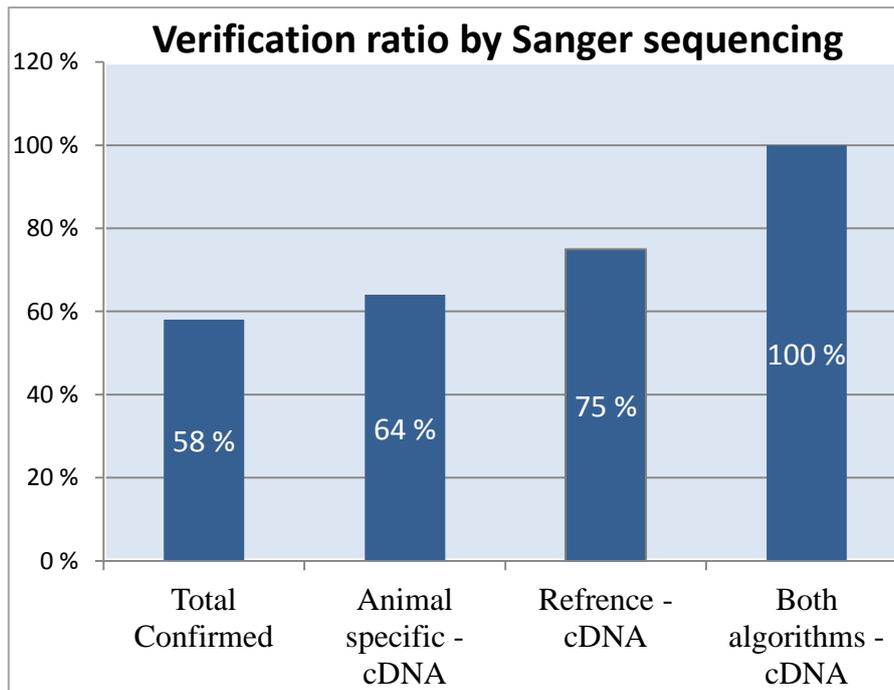


Figure 22: The diagram shows verification rate of 22 point mutations identified by exome sequencing and verified with Sanger sequencing (total confirmed). Two mutations callings, reference and animal specific call identified with cDNA sequencing and mutations found by both callings identified in cDNA is shown.

The *Kras* mutation in series C was also validated by Sanger sequenced of genomic DNA. Genomic DNA was sequenced from primary, 0<sup>th</sup> generation, 5<sup>th</sup> generation, and 10<sup>th</sup> generation (late) mammary tumors and the percentage of identified *Kras* mutation were compared with results from cDNA and exome sequencing (table 7).

Table 7: The table list *Kras* sequencing results from exome sequencing of GLI1 induced mammary tumors, Sanger sequencing of cDNA and genomic DNA, and normal liver tissue. These results are from series C from primary, 0<sup>th</sup>, 5<sup>th</sup>, and late tumors from the 10<sup>th</sup> generation transplantation, and liver from the primary mouse. It is also included if the mutations were been detected by reference call and/or animal specific call.

Gene	Sample	Exome	Animal Specific call	Reference call	cDNA	Genomic DNA
<b>Kras</b>	Primary	9	+	-	63	25
	0 Generation					42
	5 Generation					44
	Late	30	+	+	75	60
	Liver	0				0

## 5.0 Discussion

### 5.1 Methodological discussion

Every lab procedure and method has limitations including, instrumental malfunction, low dynamic range, malpractice, as well as functional limitations. Most of the methods used in this thesis are well established and optimized over years but some limitations still exist and are important to consider

#### 5.1.1 Immunohistochemistry

IHC staining is based on correctly identifying positive and negative cells for a specific marker. The major problems with IHC are high background staining and too weak specific staining. The IHC protocol includes many different reagents and it is important that they are of good quality to function properly. High background can be caused by too high concentration of the primary or secondary antibody, or active endogenous peroxidases or phosphatases. False positive staining may be due to nonspecific binding of primary antibody to identical or similar non-targeted antigens. High primary antibody concentration might increase nonspecific binding and too high secondary antibody concentration may cause binding to epitopes without primary antibody. False negative cells might be due to protein degradation and loss of the antigen or incomplete epitope retrieval<sup>156</sup>.

Based on the limitations known with IHC, it was important to include positive controls and verify that the expected cells were positive, e.g. that cytokeratins stained the expected epithelial cell layers. If a positive control could not be found within the tumor or normal tissue, a separate positive control had to be included. For each of the antibodies and slides, one of two sections was not incubated with the primary antibody, but everything else was performed in the same way. This sample worked as a negative control and as a check that the staining was only due to the primary antibody and other factors that may have caused the positive staining were eliminated. Staining of luminal and basal cells in normal mammary gland tissue was used to confirm that the cytokeratin antibodies stained the expected epithelial cells.

### 5.1.2 DNA and RNA isolation

The major source of error in DNA and RNA isolation is getting low yield and risk of contaminated samples. Contamination can be due to proteins that are not removed from the sample or other organic contaminants from the extraction procedure. Error may also occur during concentration measuring, especially with NanoDrop when the sample is not pure. RNA isolated using the TRIzol base protocol gave satisfactory results for all samples except for one, which was successfully isolated with Quick-RNA MiniPrep kit. RNA yield from Quick-RNA MiniPrep kit was lower compared to TRIzol isolated RNA but was of no concern for further analysis in this thesis.

### 5.1.3 cDNA synthesis

cDNA synthesis with incorrect temperatures during annealing and synthesis will affect the outcome of the synthesis. Too high temperature during the synthesis could make the reverse transcriptase enzyme inactive, and too low temperature the enzyme may not work optimal. In addition, the reverse transcriptase might insert wrong nucleotides resulting in false mutations in the synthesized cDNA.

cDNA concentrations were first measured on NanoDrop. Based on these concentrations the resulting PCR products in the gel electrophoresis were very weak. NanoDrop measures everything in the samples that absorb light at that wavelength, including nucleotides, and it was thereby measured 200 times higher concentration than the Qubit 2.0 fluorometer. After using the concentration obtained with the Qubit instrument, the PCR for the Sanger sequencing improved.

Measuring cDNA with Qubit dsDNA HS Assay Kit is not optimal since it recognizes dsDNA, and the samples measured contained cDNA:mRNA hybrid. However, this was the best solution to determine cDNA concentration with the equipment available. One alternative option could have been to conduct second strand synthesis to obtain cDNA:cDNA hybrid and subsequently measured these concentrations with the Qubit instrument.

#### 5.1.4 Advantages and disadvantages with Sanger and SOLiD sequencing technologies

Both SOLiD and Sanger technologies have advantages and disadvantages. SOLiD sequences up to 50 base pairs which is significantly less than Sanger sequencing that can sequence up to 900 base pairs. In Sanger sequencing up to 96 reactions can be performed at the same time, but in SOLiD up to a billion of reactions can be performed in parallel. The raw data from SOLiD are superior to Sanger, but have overall an error rate that is higher due to the difficulty of sequence assembly. SOLiD sequences deeper than Sanger which is important for detecting mutations, sequence variants or massively rearranged regions. But great depths mean more sequencing and more difficulties in de novo assembly<sup>142, 157, 158</sup>.

#### 5.1.5 Exome sequencing

The animal specific call was considered to be the most correct way of detecting mutations in the exome sequencing data. Exome sequencing results aligned with all known reference genomes were used to account for the mixed genetic background of the mice. We also aligned data with DNA sequenced from liver to remove naturally occurring mutations in the mouse.

The number of times each nucleotide has been sequenced in exome sequencing reflects the depth of the sequencing (coverage). The sequencing depth of the mutations chosen for Sanger sequencing validation varied greatly, from below 10 counts to over 300 counts for the position where the mutations are. The majority of the samples were sequenced at a depth of 70. Sequencing depth of 30 for SOLiD technology is considered to be very accurately<sup>159</sup>. The average read depth is therefore sufficient for most of the genes, but for *Atp13a3*, *Brca2* and *Brcc3* that have read depth below 15, the exome sequencing result is not as confident.

Difficulties with exome sequence assembly may have lead to incorrect alignment, and may have caused wrong identification of mutation which could lead to falsely identified or unidentified mutation.

#### 5.1.6 Sanger sequencing

Originally two additional genes were chosen to be sequenced (*Ovgp1* and *Skint5*) but it was not possible to design adequate primers for both sequencing directions. Thus, analyses of these genes are not included in the thesis. To conclude that a mutated gene is expressed, it is

necessary to confirm the mutations in cDNA by sequencing in both directions. This principle was established by Department of Genetics and followed in this thesis.

After the amplification step and separation on gel electrophoresis, showed only one band were present using the different samples and primer pairs but the bands showed variable intensities. This may be due to laboratory difficulties with samples application, non-optimal cDNA concentrating or insufficient cDNA input. Gel electrophoresis of gDNA gave the expected bands but in addition a smear was observed, probably cause by high DNA concentration and not by unspecific binding of the primers resulting in many bands of different sizes.

The signal intensity from sequencing analysis software varied throughout the sequence both in signal evenness within the same sample and in signal strength between samples. The first and last part of the sequences was often of poor quality and was not considered.

Most of the negative controls, reactions without gDNA, had signal intensities as high as the samples but no readable sequences. The negative controls were considered negative since there were no readable sequences and the signal could be considered as noise. In addition, normal mouse breast tissue was also included as negative controls which all were negative for the point mutation and did not show similar noise as the negative controls.

Many of the sequenced samples were too noisy to draw a conclusion after the first round of sequencing but resequencing gave good enough sequences for most samples. For five samples it was not possible to obtain sequences in both directions even though the primer pairs worked with other samples. Forward Il7r late tumor, forward Il7r NOD/SCID control, reverse Brca2 NOD/SCID, forward Kdm6a NOD/SCID and forward Tsc1 NOD/SCID results are therefore not available. Sequences from the other direction of these five samples were negative for the mutation and hence, the samples were considered as wild type. Sequencing of genomic DNA in series C, in the forward direction in the primary tumor and 5<sup>th</sup> generation tumor failed. The forward direction of those two samples was scheduled to be resequenced, but the laser in the 3730 DNA Analyzer instrument was defect and there was not enough time to repair the laser and resequence the gDNA samples. However, since these *Kras* mutation had been verified with cDNA sequencing, it is reasonable to assume that the mutations are present. In addition *Btnl6* were not detected in any of the samples that were sequenced. New primer pairs designed for the exon with the mutation did not either reveal any sequence, and it was concluded that the *Btnl6* gene most likely is not expressed.

Upon sequencing of Foxp1 and Sfi1 splice variants was identified. The sequence was normal up to the exon-exon border (1591-1592bp) in both directions in Foxp1 (figure 23) indicating alterative splicing occurs at this border. New primers designed for the specific exon with the mutation did not give any sequence results. Multiple sequence alignment with sequences obtained with the old primers revealed that Foxp1 sequence was expressed and the highest peaks of the electropherogram were indeed the sequence of interest. Therefore it could be concluded that the new primers did not work. Examining the obtained sequences with the old primers in the reverse direction, the primary tumor was negative for the mutations, while the late tumor displays the mutation. However the mutation has to be validated in both directions before certain identification.

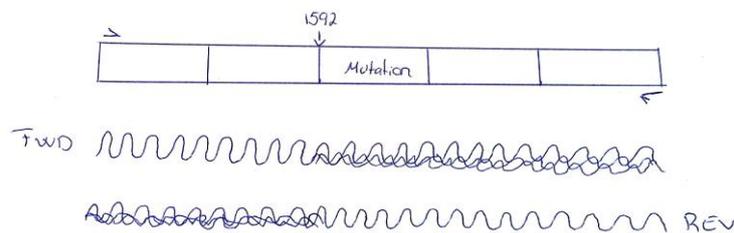


Figure 23: Sanger sequencing with Foxp1 transcript on cDNA revealed alterative splice variant, where the exon border 1591-1592 of the reference sequence are alternatively spliced

Sfi1 sequencing results showed multiple sequences in the middle of the sequenced area, indicating alternative splice variants. Comparing the sequencing results with reference sequence did not give a clear indication of the splicing site (figure 24). In both forward and reverse direction, the noise started at a different exon-exon border, and ended in the middle of an exon in both directions. The multiple sequences started at position 35 and ended at position 254, in both directions. One might suspect that the primers had some sort of unspecific binding. The gene was clearly expressed, so new primer pairs targeting only the exon of interest were designed and sequencing with one new and one old primer gave sequence results that could be used for validation.

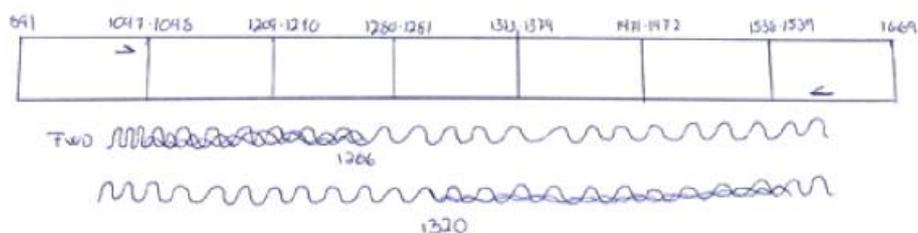


Figure 24: Sanger sequencing results of Sfi1 cDNA with old primers compared with the reference sequence

All of the readable sequences aligned to the expected area on the reference sequences. Since the beginning and end of the sequences had poorer quality the bases in those areas can potentially be incorrectly identified. This type of error was easily identified in multiple sequence alignments and was overlooked in the validation of the various mutations.

### 5.1.7 Comparing exome and Sanger sequencing results

Sanger sequencing is not considered as a quantitative method but rather a method for detecting mutations. Sanger sequencing was performed to validate exome sequencing results. In this thesis the signal strength between wild type and mutation nucleotide were used to compare with the exome sequencing result. Comparison of the percentage mutation detection by exome and Sanger sequencing showed in some cases very different results, even more than 50% difference in one case. Even comparing genomic DNA with cDNA results for the *Kras* mutation showed 30% difference. Sequencing of cDNA detected in general a higher percentage mutation than by using genomic DNA. Of the total of 31 mutations that was identified by exome sequencing in the different tumor samples and chosen for verification, 42% of the mutations were not verified by cDNA sequencing.

Tumor heterogeneity can explain the difference between the results, where one area of the tumor might have more cells with the mutation than another part of the tumor. Cells with *Kras* mutation express much more *Kras* than cells without the mutated gene, indicating a skewed relationship between mutated cells and non-mutated cells. For most of the other mutations analyzed in this thesis, there is not much information available about the consequences of the mutations as it is for *Kras* mutations. Technical aspects might also contribute to the observed differences in the validation percentage. To further improve the quality of the information concerning the mutation status in the GLI1 induced tumors, samples have been subjected to Illumina massively parallel sequencing, an alternative method to SOLiD sequencing.

## 5.2 Biological discussion

Overexpression of GLI1 alone is not sufficient to induce mammary cancer. Therefore, other factors contributing to tumor development are likely to be involved. In this study, tumors developed in a GLI1 transgenic mouse model were characterized and analyzed for gene mutations. Such mouse models are important for identification of targeted therapy for GLI1 induced mammary tumors. Whole genome and exome sequencing identified genetic alterations likely to lead to tumor development, however, these approaches are not flawless, hence validation was necessary to claim somatic mutations present in the GLI1 induced tumors.

### 5.2.1 Tumor characterization

The tumors were characterized by IHC. The GLI1 overexpressing tumors showed both luminal and basal-like characteristics. One series (series B) was heterogeneous with respect to these cell characteristics, and both luminal and basal-like cells were observed in the same tumor samples. GLI1 overexpressing tumors can therefore not be categorized as either luminal or basal-like tumors, and this finding has also been shown previously<sup>20</sup>. The GLI1 overexpressing tumors show different differentiation patterns based on H&E staining and different differentiation patterns in GLI1 induced tumors have been reported by others as well<sup>20</sup>. This showed that GLI1 tumors do not fall into one category of breast cancer types.

The GLI1 tumors were negative for ER-alpha, Her2 and PR and thereby classified as triple-negative breast cancer (TNBC), a subtype representing 12-17% of human breast tumors. Triple-negative breast cancer is almost always basal-like<sup>160</sup>. Since GLI1 overexpressing mammary tumors have both luminal and basal-like characteristics, this model does not fit in the known subtype profile. In support of this, the tumors were negative for Egfr which is unusual for basal-like breast cancer. The proliferation status of the tumors increased during the serial transplantation as measured by Ki-67, suggesting that the GLI1 overexpressing tumors develop into highly proliferative and potentially aggressive tumors and is further strengthened by high Ki-67 staining is associated with TNBC are more aggressive tumors have been reported<sup>112</sup>.

Erk1/2 is phosphorylated in the cytoplasm and translocates to the nucleus, hence cytoplasmic staining can therefore be expected as it was observed in the GLI1 induced tumors. Cytoplasmic staining has been reported by others<sup>161, 162</sup> and it was further described that

cytoplasmic stained p-Erk1/2 probably reflects only a fraction of the total p-Erk1/2<sup>162</sup>. In this thesis, the tumors had more cytoplasmic staining compared with nuclear staining. Tumors from series C and D showed some nuclear staining of p-Erk1/2 which might be explained by mutated *Kras* that increase the MAPK signaling leading to more p-Erk1/2 activation. Late tumor in series B also had a few nuclear stained cells. *Egfr* which also activates the MAPK signaling pathway were not identified in those tumors, this further strengthening that mutated *Kras* activates the MAPK signaling pathway.

### 5.2.2 Gene mutations and impact on tumorigenesis

All cancers carry somatic mutations and somatic mutations are part of normal biology. Some somatic mutations are repaired by the cell's DNA repair machinery and others will lead to tumor development<sup>18</sup>. Germline mutations and normal variants were excluded from the identified point mutations from the exome sequencing by aligning sequences from breast tumor samples with those from liver samples. Furthermore, germline mutations and variants were also excluded by the verification step using Sanger sequencing from cDNA. By using cDNA, presence of the mutated expressed transcript was ensured.

Whole exome sequencing gives insight to mutational signatures and processes<sup>163</sup>. Most mutations in breast cancer probably occurs after driver mutation initiation and an increasing number of mutations are seen with more malignant tumors<sup>18</sup>. Only a few mutations are repeatedly found in breast tumors whereas rare mutations are more common. Also, several rare mutations are often found in different combinations which contribute to tumor development. The pathways associated with cancer are more important than the mutated genes themselves, which can be grouped and associated with specific cellular processes and signaling pathways<sup>164</sup>.

The 20 genes verified by Sanger sequencing have human homologous except *Btln6* according to COSMIC database and the mutated genes have been reported in human breast cancers. Of the 20 genes, 18 were present in the animal specific call data (see appendix A). Only two of the genes, *Kras* and *Cdk4* (position G67A), have mutation detected at the same position in human and only *Kras* have the same amino acid substitution reported in human cancers. Nine of the 18 genes are present COSMIC Cancer Gene Census, meaning these genes have been reported

to causes cancer. These nine genes include *Cdk4*, *Foxp1*, *Kdm6a*, *Kras*, *Il7r*, *Maml2*, *Psip1*, *Ptch1* and *Tsc1*.

Of the 20 genes sequenced by Sanger sequencing, 10 genes were identified to have the mutation expressed in the tumors; *Atp13a3*, *Commd7*, *Fam179b*, *Kdm6a*, *Kras*, *Lpin1*, *Maml2*, *Myom1*, *Stat6* and *Tsc1* (overview of mutation in the different series, and tumor samples and verified vs. not is listed in appendix A). Only *Fam179b* in late tumor was homozygous for the mutation. These 10 genes have been reported mutated in human breast cancer according to COMSIC database, but only the specific *Kras* mutation have previously been found in other breast cancer samples. Moreover, only *Kdm6a*, *Kras*, *Maml2*, and *Tsc1* were present in COSMIC Cancer Gene Census and have mutations that are causally related to cancer development. *Maml2*, *Myom1* and *Tsc1* were gained mutations and could therefore not be considered as driver mutations in GLI1 induced tumors. All of the gained mutations were found in series C. *Kdm6a* was loosed but could still have been a driver mutation in series B. It is likely that *Kras* was a driver mutation in series C and D since no mutation could explain tumor development in these series. Finally, eight genes showed only expression wild type while two genes were not available due to malfunction primer pairs and one gene not being expressed. Series A did not have any confirmed mutations by Sanger sequencing meaning that no driver mutations were validated by the chosen genes to be Sanger sequenced in series A.

Both gained and lost mutated genes showed low expression of the mutation and it is possible that in those cases where the mutations were not identified in the tumors of the same series, the mutation may have been below detection limit. Stephens et al.<sup>18</sup> argue that most mutations in breast cancer occur after driver mutation initiation and those increasing numbers of mutations are found in more tumors, which may explain the findings of gained mutations. These results could indicate that the cells with somatic mutation compromises a higher percentage of the tumor cells in the late tumors, or that the tumor cells express more of the mutated genes compared to the cells without the mutation.

No mutations found by exome sequencing were identified in all four series and very few mutations were found in several series, indicating that several mutations are potentially involved in GLI1 induced tumor development. Not many mutations were identified in both primary and late tumors by the exome sequencing, indicating dynamic tumor progression and mutations that are gained and lost.

There are still several hundred mutations that were not selected for verification that could potentially be involved in tumor development. For the purpose of this study, mutated genes reported in COSMIC Cancer Gene Census were prioritized, but other genes not known to be involved in cancer could still be important for GLI1 induced tumors.

Another important factor is gene amplification. Gene amplifications are frequent in breast cancer. Stephans et al <sup>18</sup> have found 40 driver mutations caused by either somatic point mutations or copy number alterations. *TP53* is a driver mutation and primary mouse in series C have *TP53* mutated gene inserted in the genome and is probably a driver mutations in this series in addition to *Kras*. Since copy numbers are not available at this point in the project and only a handful of genes that have been verified, it is difficult to pinpoint driver mutations. The tumor may also have several driver mutations that could not be identified in this thesis.

Epigenetic is a reversible process and regulates gene expression and could explain why no identification was verified in series A and that epigenetic alterations may have for instance inhibited a tumors suppressor gene. Series A do not have any identified mutations and this could also imply that maybe GLI1 alone is sufficient for tumor development.

### 5.2.3 Tumor stability

For the major part, the GLI1 induced tumors showed stable characteristics throughout the serial transplantations. Cytokeratin staining varied to some extent between the serially transplanted tumors and across series, but this had no impact on classification between primary and late tumors. The hormone receptors were quite stable in the degree of staining, except for PR that showed cytoplasmic staining in the late tumors compared to no staining in the primary tumors. p-Erk1/2 showed only minor variation in the staining but the largest differences were observed for Ki-67 for which positively stained cells were observed in the late tumors versus non in the primary tumors. The large differences seen between the result from exome sequencing and Sanger sequencing warranted additional sequencing analysis to clarify these variations.

### 5.3 Further perspectives

Thorough characterization of the GLI1 oncogene induced tumors is important for their use as breast cancer models. Further investigation of these tumors is a current ongoing project. Several things could be of interest for further investigations:

- Verification of exome sequencing with cDNA sequencing with Sanger sequencing technology showed great variation and was not satisfactory. Genome sequencing with Sanger sequencing technology have been decided to perform as a third validation before concluding certain identified genetic alterations in GLI1 induced tumors. Genomic DNA was sent in the first week of May and sequencing result are available in the near future.
- Breast cancer have more often gene amplification than genetic alterations, and hopefully the genome Sanger sequencing could also be used to identify copy number variation in the mammary gland tumors
- Microarray analysis of gene expression results are currently attempted to be merged with other mouse model tumors to a cohort, to try to group the GLI1 tumors together and show that GLI1 tumors have it own mutational signature
- It would also be of interest to try and classify the GLI1 tumors according to human classifications with the use of microarray analysis
- Epigenetic could be of interest to investigate, to reveal activated or inhibited signaling pathway. Epigenetic alterations work together with genetic mutations in initiation and progression of cancer
- Sequencing with Ion torrent has the possibility to sequence much deeper than both SOLiD and Sanger sequencing. Ion torrent can sequence as deep as 30.000 reads, which can be of interest to detect low expressed mutations that were not possible to detect by Sanger sequencing. Ion torrent is not suitable for whole genome sequencing but it is possible to deep sequence genomic DNA of smaller area
- The major goal in the future is to identify signaling pathways involved in tumor development for GLI1 overexpressing tumors and develop a targeted therapy for GLI1 induced tumors
- It would also be of interest to investigate if the mouse also express their Gli1 gene and not just the human GLI1 gene to if the Hedgehog signaling pathway have been activated in the mouse models

- RNAscope, which detects mRNA in paraffin embedded tissue, was invested by the Department of Genetics were recently, and it would be of great interest to use this and look at the expression of Gli1 in the tumor cells among other expressed genes

## 6.0 Conclusion

GLI1 overexpression in the mammary gland has been shown to give rise to breast cancer. GLI1 alone is not sufficient to induce breast cancer so additional potential tumor inducing factors were investigated. Four series of GLI1 induced mammary gland tumors showed different somatic mutations and tumor characteristics. GLI1 tumors have been found to give rise to several types of breast cancer including carcinosarcoma which is a seldom breast tumor type. It was found that GLI1 induced tumors had both luminal and basal-like characteristics, and were characterized as a triple-negative subtype which is associated with basal-like tumors. However luminal tumors were also identified, showing that GLI1 induced tumors do not fit in to the usual molecular classification scheme of breast cancer. This is further supported by their lack of Egfr expression that are most often found among basal-like tumors. The GLI1 tumors showed increasing proliferation features throughout serial transplantations, indicating that the GLI1 tumors have high proliferative capacity which is associated with aggressive tumors. p-Erk1/2 was found in two of the series that also have *Kras* mutations suggesting that *Kras* have an impact on GLI1-dependent tumor development. Validation of exome sequencing identified 10 possible mutated genes that were expressed in the mammary tumors, but only *Kras* have so far been considered as a mutation that most likely has an impact on the tumor development. Further investigation is necessary to identify whether any of the additional verified mutations may have contributed to tumor development.

## Literature

1. Watson, C. J. & Khaled, W. T. (2008). Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development*, 135 (6): 995-1003.
2. Ali, S. & Coombes, R. C. (2002). Endocrine-responsive breast cancer and strategies for combating resistance. *Nat Rev Cancer*, 2 (2): 101-12.
3. World Health Organization. (2013). *Cancer*. Geneva, Switzerland: World Health Organization. Available at: <http://www.who.int/mediacentre/factsheets/fs297/en/index.html> (accessed: 29.01.2014).
4. Cancer Registry of Norway. (2013). *Cancer Statistics*. Oslo, Norway: Cancer Registry of Norway. Available at: <http://www.kreftregisteret.no/no/Registrene/Kreftstatistikk/> (accessed: 29.01.2014).
5. Cancer Research UK. (2012 ). *Breast cancer survival statistics*. Oxford, UK: Cancer Research UK. Available at: <http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/survival/breast-cancer-survival-statistics> (accessed: 13.05.2014).
6. Cancer Research UK. (2014). *Breast cancer risk factors*. Oxford, UK: Cancer Research UK Available at: <http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/riskfactors/breast-cancer-risk-factors> (accessed: 13.05.2014).
7. Hanahan, D. & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144 (5): 646-74.
8. Weniberg, R. A. (2007). *The biology of cancer*. New York, USA: Garland Science, Taylor & Francis Group, LLC.
9. Allred, D. C., Mohsin, S. K. & Fuqua, S. A. (2001). Histological and biological evolution of human premalignant breast disease. *Endocr Relat Cancer*, 8 (1): 47-61.
10. Polyak, K. (2007). Breast cancer: origins and evolution. *J Clin Invest*, 117 (11): 3155-63.
11. Bombonati, A. & Sgroi, D. C. (2011). The molecular pathology of breast cancer progression. *J Pathol*, 223 (2): 307-17.
12. Myal, Y., Leygue, E. & Blanchard, A. A. (2010). Claudin 1 in breast tumorigenesis: revelation of a possible novel "claudin high" subset of breast cancers. *J Biomed Biotechnol*, 2010: 956897.
13. Polyak, K. (2011). Heterogeneity in breast cancer. *J Clin Invest*, 121 (10): 3786-8.
14. Balmain, A., Gray, J. & Ponder, B. (2003). The genetics and genomics of cancer. *Nat Genet*, 33 Suppl: 238-44.
15. Sharma, S., Kelly, T. K. & Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis*, 31 (1): 27-36.
16. Croce, C. M. (2008). Oncogenes and cancer. *N Engl J Med*, 358 (5): 502-11.
17. Petitjean, A., Achatz, M. I., Borresen-Dale, A. L., Hainaut, P. & Olivier, M. (2007). TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene*, 26 (15): 2157-65.
18. Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G. R., et al. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature*, 486 (7403): 400-4.
19. Williams, G. H. & Stoeber, K. (2012). The cell cycle and cancer. *J Pathol*, 226 (2): 352-64.
20. Fiaschi, M., Rozell, B., Bergstrom, A. & Toftgard, R. (2009). Development of mammary tumors by conditional expression of GLI1. *Cancer Res*, 69 (11): 4810-7.

21. Downward, J. (2003). Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer*, 3 (1): 11-22.
22. Lydon, J. P., DeMayo, F. J., Funk, C. R., Mani, S. K., Hughes, A. R., Montgomery, C. A., Jr., Shyamala, G., Conneely, O. M. & O'Malley, B. W. (1995). Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes Dev*, 9 (18): 2266-78.
23. Laidlaw, I. J., Clarke, R. B., Howell, A., Owen, A. W., Potten, C. S. & Anderson, E. (1995). The proliferation of normal human breast tissue implanted into athymic nude mice is stimulated by estrogen but not progesterone. *Endocrinology*, 136 (1): 164-71.
24. Sandhu, R., Parker, J. S., Jones, W. D., Livasy, C. A. & Coleman, W. B. (2010). Microarray-Based Gene Expression Profiling for Molecular Classification of Breast Cancer and Identification of New Targets for Therapy. *Lab Medicine*, 41 (6): 364-372.
25. Nusslein-Volhard, C. & Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287 (5785): 795-801.
26. Mohler, J. (1988). Requirements for hedgehog, a segmental polarity gene, in patterning larval and adult cuticle of *Drosophila*. *Genetics*, 120 (4): 1061-72.
27. Kasper, M., Jaks, V., Fiaschi, M. & Toftgard, R. (2009). Hedgehog signalling in breast cancer. *Carcinogenesis*, 30 (6): 903-11.
28. Varjosalo, M. & Taipale, J. (2008). Hedgehog: functions and mechanisms. *Genes Dev*, 22 (18): 2454-72.
29. Hui, M., Cazet, A., Nair, R., Watkins, D. N., O'Toole, S. A. & Swarbrick, A. (2013). The Hedgehog signalling pathway in breast development, carcinogenesis and cancer therapy. *Breast Cancer Res*, 15 (2): 203.
30. Burke, R., Nellen, D., Bellotto, M., Hafen, E., Senti, K. A., Dickson, B. J. & Basler, K. (1999). Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. *Cell*, 99 (7): 803-15.
31. Oro, A. E., Higgins, K. M., Hu, Z., Bonifas, J. M., Epstein, E. H., Jr. & Scott, M. P. (1997). Basal cell carcinomas in mice overexpressing sonic hedgehog. *Science*, 276 (5313): 817-21.
32. Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., et al. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science*, 314 (5797): 268-74.
33. Xie, J., Murone, M., Luoh, S. M., Ryan, A., Gu, Q., Zhang, C., Bonifas, J. M., Lam, C. W., Hynes, M., Goddard, A., et al. (1998). Activating Smoothed mutations in sporadic basal-cell carcinoma. *Nature*, 391 (6662): 90-2.
34. Xie, J., Johnson, R. L., Zhang, X., Bare, J. W., Waldman, F. M., Cogen, P. H., Menon, A. G., Warren, R. S., Chen, L. C., Scott, M. P., et al. (1997). Mutations of the PATCHED gene in several types of sporadic extracutaneous tumors. *Cancer Res*, 57 (12): 2369-72.
35. Vorechovsky, I., Benediktsson, K. P. & Toftgard, R. (1999). The patched/hedgehog/smoothened signalling pathway in human breast cancer: no evidence for H133Y SHH, PTCH and SMO mutations. *Eur J Cancer*, 35 (5): 711-3.
36. Xu, L., Kwon, Y. J., Frolova, N., Steg, A. D., Yuan, K., Johnson, M. R., Grizzle, W. E., Desmond, R. A. & Frost, A. R. (2010). Gli1 promotes cell survival and is predictive of a poor outcome in ERalpha-negative breast cancer. *Breast Cancer Res Treat*, 123 (1): 59-71.
37. O'Toole, S. A., Machalek, D. A., Shearer, R. F., Millar, E. K., Nair, R., Schofield, P., McLeod, D., Cooper, C. L., McNeil, C. M., McFarland, A., et al. (2011). Hedgehog overexpression is associated with stromal interactions and predicts for poor outcome in breast cancer. *Cancer Res*, 71 (11): 4002-14.

38. Naylor, T. L., Greshock, J., Wang, Y., Colligon, T., Yu, Q. C., Clemmer, V., Zaks, T. Z. & Weber, B. L. (2005). High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. *Breast Cancer Res*, 7 (6): R1186-98.
39. Nessling, M., Richter, K., Schwaenen, C., Roerig, P., Wrobel, G., Wessendorf, S., Fritz, B., Bentz, M., Sinn, H. P., Radlwimmer, B., et al. (2005). Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. *Cancer Res*, 65 (2): 439-47.
40. Fiaschi, M., Rozell, B., Bergstrom, A., Toftgard, R. & Kleman, M. I. (2007). Targeted expression of GLI1 in the mammary gland disrupts pregnancy-induced maturation and causes lactation failure. *J Biol Chem*, 282 (49): 36090-101.
41. Taketani, Y. & Oka, T. (1983). Biological action of epidermal growth factor and its functional receptors in normal mammary epithelial cells. *Proc Natl Acad Sci U S A*, 80 (9): 2647-50.
42. Luetteke, N. C., Qiu, T. H., Fenton, S. E., Troyer, K. L., Riedel, R. F., Chang, A. & Lee, D. C. (1999). Targeted inactivation of the EGF and amphiregulin genes reveals distinct roles for EGF receptor ligands in mouse mammary gland development. *Development*, 126 (12): 2739-50.
43. Edery, M., Pang, K., Larson, L., Colosi, T. & Nandi, S. (1985). Epidermal growth factor receptor levels in mouse mammary glands in various physiological states. *Endocrinology*, 117 (1): 405-11.
44. Sainsbury, J. R., Farndon, J. R., Sherbet, G. V. & Harris, A. L. (1985). Epidermal-growth-factor receptors and oestrogen receptors in human breast cancer. *Lancet*, 1 (8425): 364-6.
45. Al-Kuraya, K., Schraml, P., Torhorst, J., Tapia, C., Zaharieva, B., Novotny, H., Spichtin, H., Maurer, R., Mirlacher, M., Kochli, O., et al. (2004). Prognostic relevance of gene amplifications and coamplifications in breast cancer. *Cancer Res*, 64 (23): 8534-40.
46. Klijn, J. G., Berns, P. M., Schmitz, P. I. & Foekens, J. A. (1992). The clinical significance of epidermal growth factor receptor (EGF-R) in human breast cancer: a review on 5232 patients. *Endocr Rev*, 13 (1): 3-17.
47. Ro, J., North, S. M., Gallick, G. E., Hortobagyi, G. N., Gutterman, J. U. & Blick, M. (1988). Amplified and overexpressed epidermal growth factor receptor gene in uncultured primary human breast carcinoma. *Cancer Res*, 48 (1): 161-4.
48. Rakha, E. A., El-Sayed, M. E., Green, A. R., Lee, A. H., Robertson, J. F. & Ellis, I. O. (2007). Prognostic markers in triple-negative breast cancer. *Cancer*, 109 (1): 25-32.
49. Cheang, M. C., Voduc, D., Bajdik, C., Leung, S., McKinney, S., Chia, S. K., Perou, C. M. & Nielsen, T. O. (2008). Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin Cancer Res*, 14 (5): 1368-76.
50. Sainsbury, J. R., Farndon, J. R., Needham, G. K., Malcolm, A. J. & Harris, A. L. (1987). Epidermal-growth-factor receptor status as predictor of early recurrence of and death from breast cancer. *Lancet*, 1 (8547): 1398-402.
51. Garrett, T. P., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Kofler, M., Jorissen, R. N., Nice, E. C., Burgess, A. W., et al. (2003). The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors. *Mol Cell*, 11 (2): 495-505.
52. De Potter, C. R., Schelfhout, A. M., Verbeeck, P., Lakhani, S. R., Brunken, R., Schroeter, C. A., Van den Tweel, J. G., Schauer, A. J. & Sloane, J. P. (1995). neu

- overexpression correlates with extent of disease in large cell ductal carcinoma in situ of the breast. *Hum Pathol*, 26 (6): 601-6.
53. Darcy, K. M., Zangani, D., Wohlhueter, A. L., Huang, R. Y., Vaughan, M. M., Russell, J. A. & Ip, M. M. (2000). Changes in ErbB2 (her-2/neu), ErbB3, and ErbB4 during growth, differentiation, and apoptosis of normal rat mammary epithelial cells. *J Histochem Cytochem*, 48 (1): 63-80.
  54. Pauletti, G., Godolphin, W., Press, M. F. & Slamon, D. J. (1996). Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. *Oncogene*, 13 (1): 63-72.
  55. Bose, R., Kavuri, S. M., Searleman, A. C., Shen, W., Shen, D., Koboldt, D. C., Monsey, J., Goel, N., Aronson, A. B., Li, S., et al. (2013). Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer Discov*, 3 (2): 224-37.
  56. Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A. & McGuire, W. L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science*, 235 (4785): 177-82.
  57. Nahta, R. (2012). Molecular Mechanisms of Trastuzumab-Based Treatment in HER2-Overexpressing Breast Cancer. *ISRN Oncol*, 2012: 428062.
  58. Owens, M. A., Horten, B. C. & Da Silva, M. M. (2004). HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. *Clin Breast Cancer*, 5 (1): 63-9.
  59. von Lintig, F. C., Dreilinger, A. D., Varki, N. M., Wallace, A. M., Casteel, D. E. & Boss, G. R. (2000). Ras activation in human breast cancer. *Breast Cancer Res Treat*, 62 (1): 51-62.
  60. Adjei, A. A. (2001). Blocking oncogenic Ras signaling for cancer therapy. *J Natl Cancer Inst*, 93 (14): 1062-74.
  61. Clark, G. J. & Der, C. J. (1995). Aberrant function of the Ras signal transduction pathway in human breast cancer. *Breast Cancer Res Treat*, 35 (1): 133-44.
  62. Roberts, P. J. & Der, C. J. (2007). Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*, 26 (22): 3291-310.
  63. Gee, J. M., Robertson, J. F., Ellis, I. O. & Nicholson, R. I. (2001). Phosphorylation of ERK1/2 mitogen-activated protein kinase is associated with poor response to anti-hormonal therapy and decreased patient survival in clinical breast cancer. *Int J Cancer*, 95 (4): 247-54.
  64. Meloche, S. (1995). Cell cycle reentry of mammalian fibroblasts is accompanied by the sustained activation of p44mapk and p42mapk isoforms in the G1 phase and their inactivation at the G1/S transition. *J Cell Physiol*, 163 (3): 577-88.
  65. Servant, M. J., Giasson, E. & Meloche, S. (1996). Inhibition of growth factor-induced protein synthesis by a selective MEK inhibitor in aortic smooth muscle cells. *J Biol Chem*, 271 (27): 16047-52.
  66. Pinhel, I. F., Macneill, F. A., Hills, M. J., Salter, J., Detre, S., A'Hern, R., Nerurkar, A., Osin, P., Smith, I. E. & Dowsett, M. (2010). Extreme loss of immunoreactive p-Akt and p-Erk1/2 during routine fixation of primary breast cancer. *Breast Cancer Res*, 12 (5): R76.
  67. Ma, L., Lan, F., Zheng, Z., Xie, F., Wang, L., Liu, W., Han, J., Zheng, F., Xie, Y. & Huang, Q. (2012). Epidermal growth factor (EGF) and interleukin (IL)-1beta synergistically promote ERK1/2-mediated invasive breast ductal cancer cell migration and invasion. *Mol Cancer*, 11: 79.
  68. Clarke, R. B., Howell, A., Potten, C. S. & Anderson, E. (1997). Dissociation between steroid receptor expression and cell proliferation in the human breast. *Cancer Res*, 57 (22): 4987-91.

69. Stoica, G. E., Franke, T. F., Moroni, M., Mueller, S., Morgan, E., Iann, M. C., Winder, A. D., Reiter, R., Wellstein, A., Martin, M. B., et al. (2003). Effect of estradiol on estrogen receptor-alpha gene expression and activity can be modulated by the ErbB2/PI 3-K/Akt pathway. *Oncogene*, 22 (39): 7998-8011.
70. O'Malley, B. W., Sherman, M. R. & Toft, D. O. (1970). Progesterone "receptors" in the cytoplasm and nucleus of chick oviduct target tissue. *Proc Natl Acad Sci U S A*, 67 (2): 501-8.
71. Conneely, O. M., Jericevic, B. M. & Lydon, J. P. (2003). Progesterone receptors in mammary gland development and tumorigenesis. *J Mammary Gland Biol Neoplasia*, 8 (2): 205-14.
72. Welshons, W. V., Lieberman, M. E. & Gorski, J. (1984). Nuclear localization of unoccupied oestrogen receptors. *Nature*, 307 (5953): 747-9.
73. Hall, J. M. & McDonnell, D. P. (1999). The estrogen receptor beta-isoform (ERbeta) of the human estrogen receptor modulates ERalpha transcriptional activity and is a key regulator of the cellular response to estrogens and antiestrogens. *Endocrinology*, 140 (12): 5566-78.
74. Speirs, V., Skliris, G. P., Burdall, S. E. & Carder, P. J. (2002). Distinct expression patterns of ER $\alpha$  and ER $\beta$  in normal human mammary gland. *Journal of Clinical Pathology*, 55 (5): 371-374.
75. Krege, J. H., Hodgin, J. B., Couse, J. F., Enmark, E., Warner, M., Mahler, J. F., Sar, M., Korach, K. S., Gustafsson, J. A. & Smithies, O. (1998). Generation and reproductive phenotypes of mice lacking estrogen receptor beta. *Proc Natl Acad Sci U S A*, 95 (26): 15677-82.
76. Mallepell, S., Krust, A., Chambon, P. & Brisken, C. (2006). Paracrine signaling through the epithelial estrogen receptor alpha is required for proliferation and morphogenesis in the mammary gland. *Proc Natl Acad Sci U S A*, 103 (7): 2196-201.
77. Iwao, K., Miyoshi, Y., Egawa, C., Ikeda, N. & Noguchi, S. (2000). Quantitative analysis of estrogen receptor-beta mRNA and its variants in human breast cancers. *Int J Cancer*, 88 (5): 733-6.
78. Roger, P., Sahla, M. E., Makela, S., Gustafsson, J. A., Baldet, P. & Rochefort, H. (2001). Decreased expression of estrogen receptor beta protein in proliferative preinvasive mammary tumors. *Cancer Res*, 61 (6): 2537-41.
79. Guiochon-Mantel, A., Loosfelt, H., Lescop, P., Christin-Maitre, S., Perrot-Applanat, M. & Milgrom, E. (1992). Mechanisms of nuclear localization of the progesterone receptor. *J Steroid Biochem Mol Biol*, 41 (3-8): 209-15.
80. Meyer, M. E., Quirin-Stricker, C., Lerouge, T., Bocquel, M. T. & Gronemeyer, H. (1992). A limiting factor mediates the differential activation of promoters by the human progesterone receptor isoforms. *J Biol Chem*, 267 (15): 10882-7.
81. Wen, D. X., Xu, Y. F., Mais, D. E., Goldman, M. E. & McDonnell, D. P. (1994). The A and B isoforms of the human progesterone receptor operate through distinct signaling pathways within target cells. *Mol Cell Biol*, 14 (12): 8356-64.
82. Silberstein, G. B., Van Horn, K., Shyamala, G. & Daniel, C. W. (1996). Progesterone receptors in the mouse mammary duct: distribution and developmental regulation. *Cell Growth Differ*, 7 (7): 945-52.
83. Brisken, C., Park, S., Vass, T., Lydon, J. P., O'Malley, B. W. & Weinberg, R. A. (1998). A paracrine role for the epithelial progesterone receptor in mammary gland development. *Proc Natl Acad Sci U S A*, 95 (9): 5076-81.
84. Sivaraman, L., Hilsenbeck, S. G., Zhong, L., Gay, J., Conneely, O. M., Medina, D. & O'Malley, B. W. (2001). Early exposure of the rat mammary gland to estrogen and

- progesterone blocks co-localization of estrogen receptor expression and proliferation. *J Endocrinol*, 171 (1): 75-83.
85. Breastcancer.org. (2014). *How to Read Hormone Receptor Test Results* Ardmore, PA, USA: Breastcancer.org Available at: [http://www.breastcancer.org/symptoms/diagnosis/hormone\\_status/read\\_results](http://www.breastcancer.org/symptoms/diagnosis/hormone_status/read_results) (accessed: 19.04.2014).
  86. Dunnwald, L. K., Rossing, M. A. & Li, C. I. (2007). Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res*, 9 (1): R6.
  87. Aaltomaa, S., Lipponen, P., Eskelinen, M., Kosma, V. M., Marin, S., Alhava, E. & Syrjanen, K. (1991). Hormone receptors as prognostic factors in female breast cancer. *Ann Med*, 23 (6): 643-8.
  88. Alberts, B., Johnson, A., Lewia, J., Raff, M., Roberts, K. & Walter, P. (2008). *Molecular Biology of The Cell*. 5th ed. New York: Garland Science.
  89. Moll, R., Franke, W. W., Schiller, D. L., Geiger, B. & Krepler, R. (1982). The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *Cell*, 31 (1): 11-24.
  90. Schweizer, J., Bowden, P. E., Coulombe, P. A., Langbein, L., Lane, E. B., Magin, T. M., Maltais, L., Omary, M. B., Parry, D. A., Rogers, M. A., et al. (2006). New consensus nomenclature for mammalian keratins. *J Cell Biol*, 174 (2): 169-74.
  91. Franke, W. W., Schmid, E., Osborn, M. & Weber, K. (1979). Intermediate-sized filaments of human endothelial cells. *J Cell Biol*, 81 (3): 570-80.
  92. Taylor-Papadimitriou, J., Stampfer, M., Bartek, J., Lewis, A., Boshell, M., Lane, E. B. & Leigh, I. M. (1989). Keratin expression in human mammary epithelial cells cultured from normal and malignant tissue: relation to in vivo phenotypes and influence of medium. *J Cell Sci*, 94 ( Pt 3): 403-13.
  93. Asch, B. B., Burstein, N. A., Vidrich, A. & Sun, T. T. (1981). Identification of mouse mammary epithelial cells by immunofluorescence with rabbit and guinea pig antikeratin antisera. *Proc Natl Acad Sci U S A*, 78 (9): 5643-7.
  94. Abd El-Rehim, D. M., Pinder, S. E., Paish, C. E., Bell, J., Blamey, R. W., Robertson, J. F., Nicholson, R. I. & Ellis, I. O. (2004). Expression of luminal and basal cytokeratins in human breast carcinoma. *J Pathol*, 203 (2): 661-71.
  95. Axlund, S. D., Yoo, B. H., Rosen, R. B., Schaack, J., Kabos, P., Labarbera, D. V. & Sartorius, C. A. (2013). Progesterone-inducible cytokeratin 5-positive cells in luminal breast cancer exhibit progenitor properties. *Horm Cancer*, 4 (1): 36-49.
  96. Nagle, R. B., Bocker, W., Davis, J. R., Heid, H. W., Kaufmann, M., Lucas, D. O. & Jarasch, E. D. (1986). Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells. *J Histochem Cytochem*, 34 (7): 869-81.
  97. Wetzels, R. H., Kuijpers, H. J., Lane, E. B., Leigh, I. M., Troyanovsky, S. M., Holland, R., van Haelst, U. J. & Ramaekers, F. C. (1991). Basal cell-specific and hyperproliferation-related keratins in human breast cancer. *Am J Pathol*, 138 (3): 751-63.
  98. Boecker, W. & Buerger, H. (2003). Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept. *Cell Prolif*, 36 Suppl 1: 73-84.
  99. Lim, E., Vaillant, F., Wu, D., Forrest, N. C., Pal, B., Hart, A. H., Asselin-Labat, M.-L., Gyorki, D. E., Ward, T., Partanen, A., et al. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med*, 15 (8): 907-913.

100. Bocker, W., Moll, R., Poremba, C., Holland, R., Van Diest, P. J., Dervan, P., Burger, H., Wai, D., Ina Diallo, R., Brandt, B., et al. (2002). Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: a new cell biological concept. *Lab Invest*, 82 (6): 737-46.
101. Woodcock-Mitchell, J., Eichner, R., Nelson, W. G. & Sun, T. T. (1982). Immunolocalization of keratin polypeptides in human epidermis using monoclonal antibodies. *J Cell Biol*, 95 (2 Pt 1): 580-8.
102. Grimm, S., Bu, W., Longley, M., Roop, D., Li, Y. & Rosen, J. (2006). Keratin 6 is not essential for mammary gland development. *Breast Cancer Research*, 8 (3): R29.
103. Eerola, H., Heinonen, M., Heikkila, P., Kilpivaara, O., Tamminen, A., Aittomaki, K., Blomqvist, C., Ristimaki, A. & Nevanlinna, H. (2008). Basal cytokeratins in breast tumours among BRCA1, BRCA2 and mutation-negative breast cancer families. *Breast Cancer Res*, 10 (1): R17.
104. Sutton, L. M., Han, J. S., Molberg, K. H., Sarode, V. R., Cao, D., Rakheja, D., Sailors, J. & Peng, Y. (2010). Intratumoral expression level of epidermal growth factor receptor and cytokeratin 5/6 is significantly associated with nodal and distant metastases in patients with basal-like triple-negative breast carcinoma. *Am J Clin Pathol*, 134 (5): 782-7.
105. Alshareeda, A. T., Soria, D., Garibaldi, J. M., Rakha, E., Nolan, C., Ellis, I. O. & Green, A. R. (2013). Characteristics of basal cytokeratin expression in breast cancer. *Breast Cancer Res Treat*, 139 (1): 23-37.
106. Lane, E. B. (1982). Monoclonal antibodies provide specific intramolecular markers for the study of epithelial tonofilament organization. *J Cell Biol*, 92 (3): 665-73.
107. Schaller, G., Fuchs, I., Pritze, W., Ebert, A., Herbst, H., Pantel, K., Weitzel, H. & Lengyel, E. (1996). Elevated keratin 18 protein expression indicates a favorable prognosis in patients with breast cancer. *Clin Cancer Res*, 2 (11): 1879-85.
108. Heatley, M., Maxwell, P., Whiteside, C. & Toner, P. (1995). Cytokeratin intermediate filament expression in benign and malignant breast disease. *J Clin Pathol*, 48 (1): 26-32.
109. Goddard, M. J., Wilson, B. & Grant, J. W. (1991). Comparison of commercially available cytokeratin antibodies in normal and neoplastic adult epithelial and non-epithelial tissues. *J Clin Pathol*, 44 (8): 660-3.
110. Bullwinkel, J., Baron-Luhr, B., Ludemann, A., Wohlenberg, C., Gerdes, J. & Scholzen, T. (2006). Ki-67 protein is associated with ribosomal RNA transcription in quiescent and proliferating cells. *J Cell Physiol*, 206 (3): 624-35.
111. Gerdes, J., Schwab, U., Lemke, H. & Stein, H. (1983). Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. *Int J Cancer*, 31 (1): 13-20.
112. Keam, B., Im, S. A., Lee, K. H., Han, S. W., Oh, D. Y., Kim, J. H., Lee, S. H., Han, W., Kim, D. W., Kim, T. Y., et al. (2011). Ki-67 can be used for further classification of triple negative breast cancer into two subtypes with different response and prognosis. *Breast Cancer Res*, 13 (2): R22.
113. Molino, A., Micciolo, R., Turazza, M., Bonetti, F., Piubello, Q., Bonetti, A., Nortilli, R., Pelosi, G. & Cetto, G. L. (1997). Ki-67 immunostaining in 322 primary breast cancers: associations with clinical and pathological variables and prognosis. *Int J Cancer*, 74 (4): 433-7.
114. Inwald, E. C., Klinkhammer-Schalke, M., Hofstadter, F., Zeman, F., Koller, M., Gerstenhauer, M. & Ortmann, O. (2013). Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. *Breast Cancer Res Treat*, 139 (2): 539-52.

115. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. (2000). Molecular portraits of human breast tumours. *Nature*, 406 (6797): 747-52.
116. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*, 98 (19): 10869-74.
117. Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*, 100 (14): 8418-23.
118. Leong, A. S. & Zhuang, Z. (2011). The changing role of pathology in breast cancer diagnosis and treatment. *Pathobiology*, 78 (2): 99-114.
119. Sorlie, T. (2004). Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. *Eur J Cancer*, 40 (18): 2667-75.
120. Herschkowitz, J. I., Simin, K., Weigman, V. J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K. E., Jones, L. P., Assefnia, S., Chandrasekharan, S., et al. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol*, 8 (5): R76.
121. Perou, C. M. & Borresen-Dale, A. L. (2011). Systems biology and genomics of breast cancer. *Cold Spring Harb Perspect Biol*, 3 (2).
122. Goldhirsch, A., Wood, W. C., Coates, A. S., Gelber, R. D., Thurlimann, B. & Senn, H. J. (2011). Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol*, 22 (8): 1736-47.
123. The Jackson Laboratory. (n.y). *Advantages of the mouse as a model organism*. Bar Harbor, Maine, USA: The Jackson Laboratory. Available at: <http://research.jax.org/mousegenetics/advantages/advantages-of-mouse.html> (accessed: 05.05.2014).
124. Pfefferle, A. D., Herschkowitz, J. I., Usary, J., Harrell, J. C., Spike, B. T., Adams, J. R., Torres-Arzayus, M. I., Brown, M., Egan, S. E., Wahl, G. M., et al. (2013). Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biol*, 14 (11): R125.
125. Helse- og omsorgsdepartementet. (2001). *8.1.3 Dyrevernløven*. Oslo, Norway: Helse- og omsorgsdepartementet. Available at: <http://www.regjeringen.no/nn/dep/hod/dok/nouer/2001/nou-2001-18/9/1/3.html?id=364430> (accessed: 05.05.2014).
126. Føllesdal. (2000). *Dyreetik*. Poland: Fagbokforlaget.
127. Lovdata. (2010). *Forskrift om forsøk med dyr*. Oslo, Norway: Lovdata. Available at: <http://lovdata.no/dokument/SF/forskrift/1996-01-15-23?q=dyrefors%C3%B8k> (accessed: 05.05.2014).
128. The Jackson Laboratory. (n.y). *Early history of mouse genetics*. Bar Harbor, Maine, USA: The Jackson Laboratory. Available at: <http://research.jax.org/mousegenetics/development/history.html> (accessed: 05.05.2014).
129. Hennighausen, L. (2000). Mouse models for breast cancer. *Breast Cancer Res*, 2 (1): 2-7.
130. SAGE Labs. (2014). *C57BL/6 Inbred Mouse*. St. Louis, MO, USA: SAGE Labs. Available at: <http://www.sageresearchlabs.com/research-models/inbred-mice/c57bl6> (accessed: 05.05.2014).

131. Davie, S. A., Maglione, J. E., Manner, C. K., Young, D., Cardiff, R. D., MacLeod, C. L. & Ellies, L. G. (2007). Effects of FVB/NJ and C57Bl/6J strain backgrounds on mammary tumor phenotype in inducible nitric oxide synthase deficient mice. *Transgenic Res*, 16 (2): 193-201.
132. Taketo, M., Schroeder, A. C., Mobraaten, L. E., Gunning, K. B., Hanten, G., Fox, R. R., Roderick, T. H., Stewart, C. L., Lilly, F., Hansen, C. T., et al. (1991). FVB/N: an inbred mouse strain preferable for transgenic analyses. *Proc Natl Acad Sci U S A*, 88 (6): 2065-9.
133. The Jackson Laboratory. (1993). *The Severe Combined Immunodeficiency (scid) Mutation*. Bar Harbor, Maine, USA: The Jackson Laboratory. Available at: <http://jaxmice.jax.org/jaxnotes/archive/453a.html> (accessed: 05.05.2014).
134. Gossen, M. & Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A*, 89 (12): 5547-51.
135. Wissozky, N. (1877). Ueber das Eosin als Reagens auf Hämoglobin und die Bildung von Blutgefäßen und Blutkörperchen bei Säugethier- und Hühnerembryonen. *Archiv für mikroskopische Anatomie*, 13 (1): 479-496.
136. Coons, A. H., Creech, H. J. & Jones, R. N. (1941). Immunological Properties of an Antibody Containing a Fluorescent Group. *Experimental Biology and Medicine*, 47 (2): 200-202.
137. KPL. (n.y). *DAB Reagent Set*. Gaithersburg MD, USA: KPL, Inc. Available at: <https://www.kpl.com/docs/datasheet/541000.PDF> (accessed: 01.04.2014).
138. Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74 (12): 5463-7.
139. Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*, 9: 387-402.
140. Applied Biosystems. (n.y). *Overview of SOLiD™ Sequencing Chemistry*. Foster City, CA, USA: Applied Biosystems. Available at: <http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-sequencing-chemistry.html> (accessed: 10.05.2014).
141. Illumina. (2013). *An Introduction to Next-Generation Sequencing Technology*. San Diego, CA, USA Illumina. Available at: [http://res.illumina.com/documents/products/illumina\\_sequencing\\_introduction.pdf](http://res.illumina.com/documents/products/illumina_sequencing_introduction.pdf) (accessed: 27.04.2014).
142. Tucker, T., Marra, M. & Friedman, J. M. (2009). Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet*, 85 (2): 142-54.
143. Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet*, 11 (1): 31-46.
144. Thermo Fisher Scientific. (2010). *Nucleic Acid*. Wilmington, DE, USA: Thermo Fisher Scientific. Available at: <http://www.nanodrop.com/ND1/NucleicAcid-Booklet.html> (accessed: 17.03.2014).
145. Molecular Probes. (2010). *Qubit™ dsDNA HS Assay Kits*. Eugene, OR: Molecular Probes, Inc. Available at: <http://tools.lifetechnologies.com/content/sfs/manuals/mp32851.pdf> (accessed: 31.01.2014).
146. Life Technologies. (2010). *Qubit 2.0 Fluorometer User Manual*. Carlsbad, CA, USA: Life Technologies Corporation. Available at: <http://tools.lifetechnologies.com/content/sfs/manuals/mp32866.pdf> (accessed: 31.03.2013).

147. Life Technologies. (n.y). *Nucleic Acid Quantitation in Solution—Section 8.3*. Carlsbad, CA, USA: Life Technologies Corporation. Available at: <http://www.lifetechnologies.com/no/en/home/references/molecular-probes-the-handbook/nucleic-acid-detection-and-genomics-technology/nucleic-acid-detection-and-quantitation-in-solution.html#head1> (accessed: 03.02.2014).
148. Vector Laboratories. (2013). *Vector M.O.M. Immunodetection Kit*. Burlingame, CA: Vector Laboratories, Inc. Available at: <https://www.vectorlabs.com/data/protocols/BMK-2202.pdf> (accessed: 16.02.2014).
149. ZYMO RESEARCH. (n.y). *Instruction manual, Quick-RNA™ MiniPrep*. ZYMO RESEARCH: ZYMO RESEARCH CORP. Available at: <http://www.zymoresearch.com/downloads/dl/file/id/152/r1054i.pdf> (accessed: 06.02.2014).
150. Thermo Fisher Scientific. (2008). *NanoDrop 1000 Spectrophotometer V3.7 User's Manual*. Wilmington, DE, USA: Thermo Fisher Scientific Inc. Available at: <http://www.nanodrop.com/library/nd-1000-v3.7-users-manual-8.5x11.pdf> (accessed: 06.02.2014).
151. Invitrogen. (n.y). *SuperScript® III Reverse Transcriptase* Life Technologies: Life Technologies. Available at: <https://www.lifetechnologies.com/order/catalog/product/18080044?ICID=search-18080044> (accessed: 30.01.2014).
152. Invitrogen. (2004). *SuperScript™ III Reverse Transcriptase*. Carlsbad, California: Invitrogen Corporation. Available at: [http://tools.lifetechnologies.com/content/sfs/manuals/superscriptIII\\_man.pdf](http://tools.lifetechnologies.com/content/sfs/manuals/superscriptIII_man.pdf) (accessed: 30.01.2014).
153. Life Technologies. (2010). *Qubit™ Fluorometric Quantitation—FAQs*. Carlsbad, CA, USA: Life Technologies Corporation. Available at: <http://www.lifetechnologies.com/content/dam/LifeTech/migration/en/filelibrary/cell-tissue-analysis/qubit-all-file-types.par.39587.file.dat/co16182-qubit-2-faq.pdf> (accessed: 31.01.2014).
154. Life Technologies. (2011). *BigDye Direct Cycle Sequencing Kit Protocol*. Carlsbad, CA, USA: Life Technologies Corporation. Available at: [http://tools.lifetechnologies.com/content/sfs/manuals/cms\\_091370.pdf](http://tools.lifetechnologies.com/content/sfs/manuals/cms_091370.pdf) (accessed: 13.03.2014).
155. Applied Biosystems. (2010). *Applied Biosystems 3730/3730xl DNA Analyzers User Guide*. Foster City, California, USA.: Applied Biosystems. Available at: [http://tools.lifetechnologies.com/content/sfs/manuals/cms\\_041259.pdf](http://tools.lifetechnologies.com/content/sfs/manuals/cms_041259.pdf) (accessed: 13.03.2014).
156. Snider, J. (n.y). *IHC Troubleshooting Guide*. Rockford, IL USA Thermo Fisher Scientific Available at: <http://www.piercenet.com/method/ihc-troubleshooting-guide> (accessed: 30.04.2014).
157. Morozova, O. & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92 (5): 255-64.
158. McKernan, K. J., Peckham, H. E., Costa, G. L., McLaughlin, S. F., Fu, Y., Tsung, E. F., Clouser, C. R., Duncan, C., Ichikawa, J. K., Lee, C. C., et al. (2009). Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res*, 19 (9): 1527-41.
159. Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol*, 2012: 251364.

160. Foulkes, W. D., Smith, I. E. & Reis-Filho, J. S. (2010). Triple-negative breast cancer. *N Engl J Med*, 363 (20): 1938-48.
161. Kuo, H. T., Hsu, H. T., Chang, C. C., Jiang, M. C., Yeh, C. M., Shen, K. H., Hsu, P. C. & Tai, C. J. (2013). High nuclear phosphorylated extracellular signal-regulated kinase expression associated with poor differentiation, larger tumor size, and an advanced stage of breast cancer. *Pol J Pathol*, 64 (3): 163-9.
162. Milde-Langosch, K., Bamberger, A. M., Rieck, G., Grund, D., Hemminger, G., Muller, V. & Loning, T. (2005). Expression and prognostic relevance of activated extracellular-regulated kinases (ERK1/2) in breast cancer. *Br J Cancer*, 92 (12): 2206-15.
163. Alexandrov, L. B. & Stratton, M. R. (2014). Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev*, 24c: 52-60.
164. Baudot, A., de la Torre, V. & Valencia, A. (2010). Mutated genes, pathways and processes in tumours. *EMBO Rep*, 11 (10): 805-10.

# Appendix

## Appendix A

Table 1: This table shows the mutated genes identified by exome sequencing and chosen to be verified on cDNA sequencing with Sanger sequencing technology in four series of GLI1 induced mammary tumor tissue from primary tumors and late tumors from the 10<sup>th</sup> generation transplantation. The gray areas represent mutations that not have been identified by exome sequencing and thereby not either been sequenced with cDNA. The yellow squares represent mutation detected by exome sequencing, either by the animal specific call (indicated by A) and/or by the reference call (indicated by R). The white squares are unavailable results due to failure in the cDNA sequencing, either by malfunction primers or that the gene is not expressed, like with Btl6 gene. The light blue squares are control sequenced samples from a normal NOD/SCID mammary gland sample. The green boxes indicates that mutation have been verified by Sanger sequencing. The Red boxes indicate that the mutations have not been verified by Sanger sequencing and only wild tumor is present. The percentage represents the amount of mutation that has been identified from the total sequencing results. The percentage detected mutation from exome sequencing is listed in table 6 in the results. The mutations identified in COSMIC Cancer Gene Census are genes that are known to cause cancer.

Genes	Series	A primary		A late		B primary		B late		C primary		C late		D primary		D late		Control, NOD/SCID
	COSMIC Cancer Gene Census	Predicted mutation	Mutation cDNA verified															
Atp13a3	-					A+R	67%	A+R	50%									
Brca2	+															R		-
Brc3	-											R						
Btl6	-													A+R		A+R		
Cdk4 G151A	+							A										-
Cdk4 G67A	+			A														-
Commd7	-												A	33%	A+R	35%		
Fam179b	-					A+R	86%	A+R	100%									
Foxp1	+											A						
Il7r	+			A														-
Kdm6a	+					A	11%											-
Kras	+									A	63%	A+R	75%	A+R	78%	A+R	88%	-
Lpin1	-					A	53%	A+R	90%									
Maml2	+											A	8%					-
Myom1	-									A		A	13%					-
Psip1	+															A		-
Ptch1 G265T	+											A						-
Ptch1 C888A	+	A																-
Sema6d	-					A		A										
Sfi1	-													A		A		
Stat6	-									A	21%	A+R	42%					
Tsc1	+											A	14%					-

# Appendix B

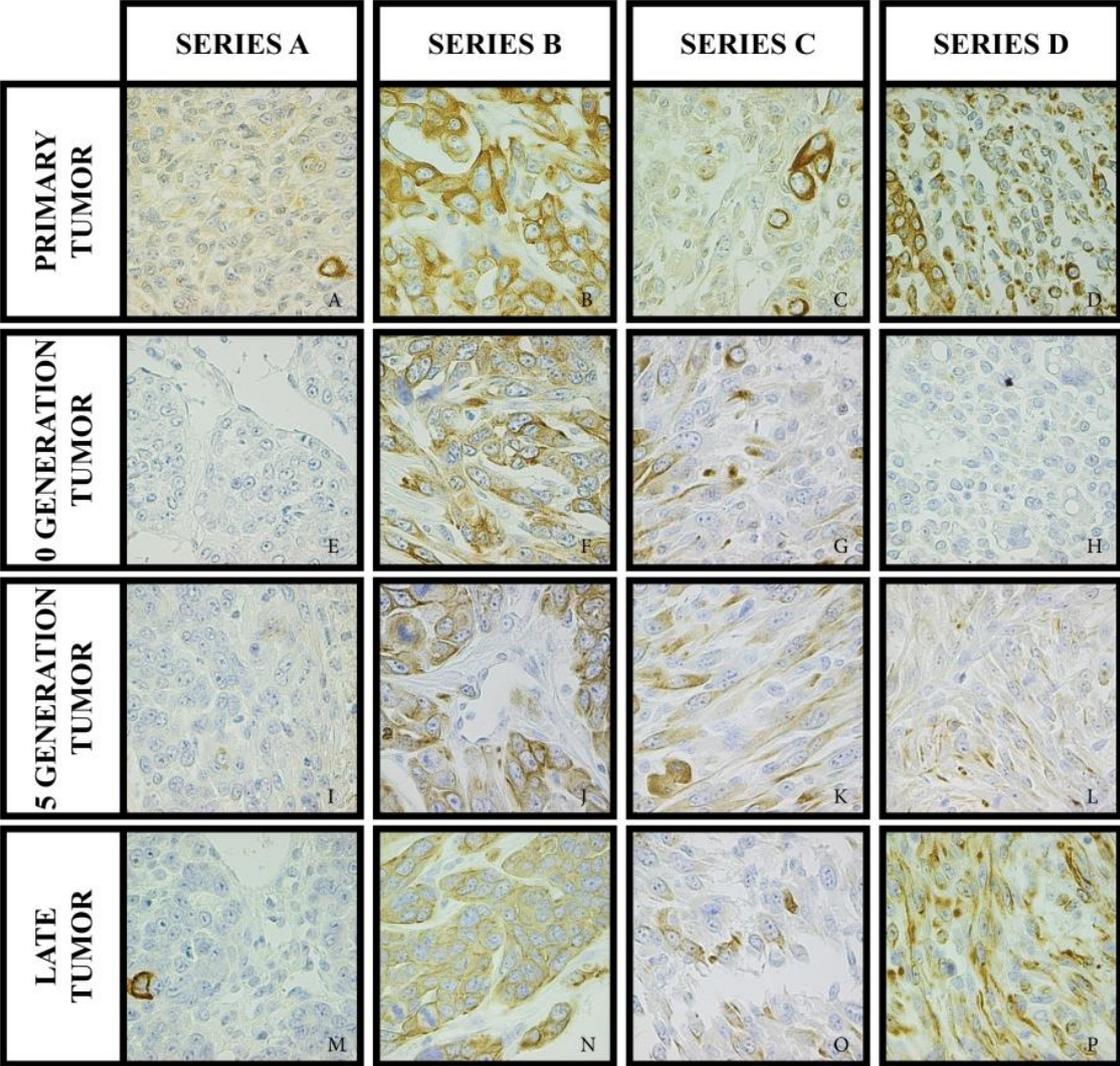


Figure 1: Four series of GLI1 induced mammary tumors stained with K18 (A-P) antibody from primary tumors (A-D), 0<sup>th</sup> (E-H), 5<sup>th</sup> (I-L) and late tumors from the 10<sup>th</sup> generation transplantation (M-P). Magnification 400x.

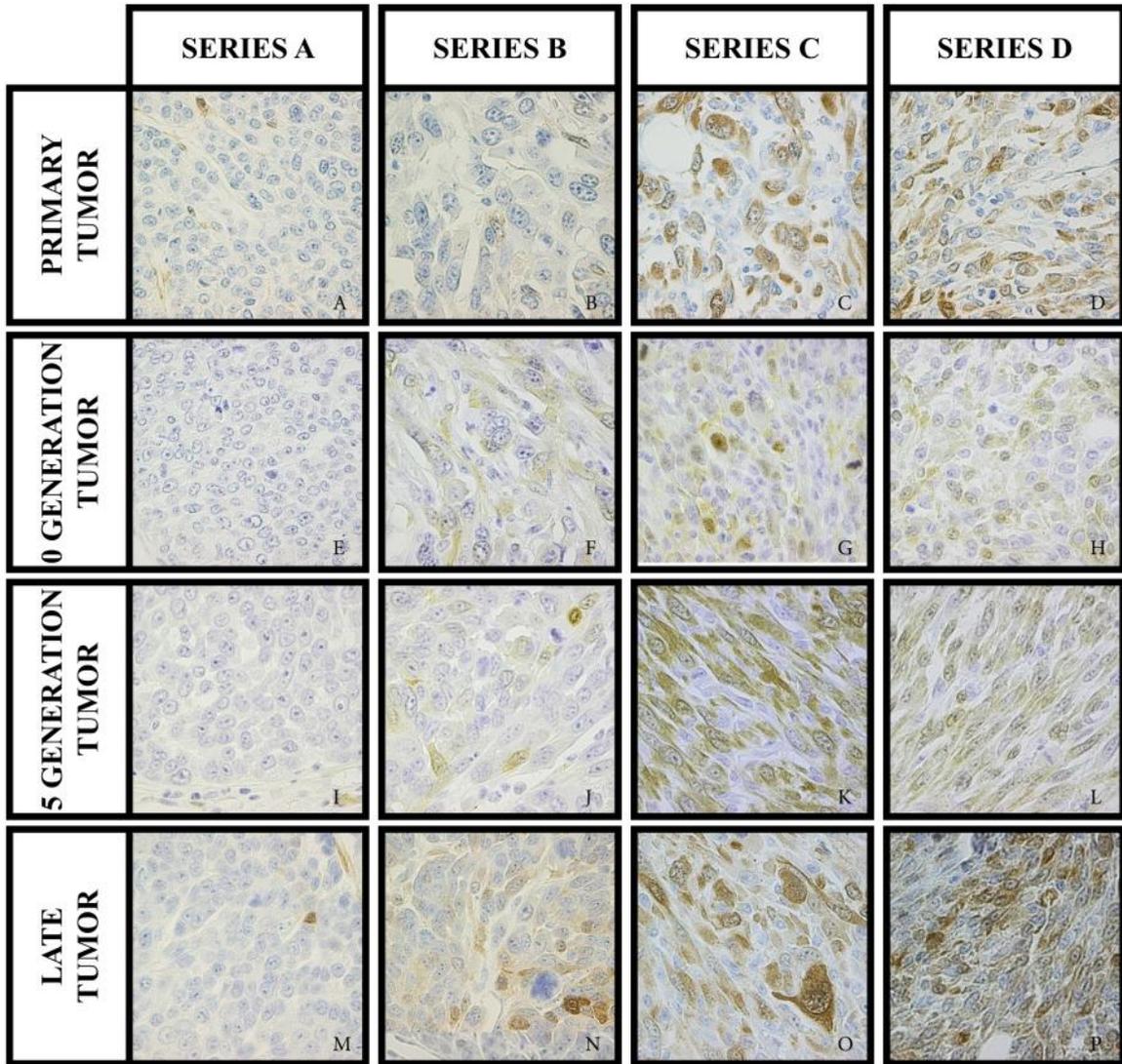


Figure 2: Four series of GLI1 induced mammary tumors stained with p-Erk1/2 (A-P) antibody from primary tumors (A-D), 0<sup>th</sup> (E-H), 5<sup>th</sup> (I-L) and late tumors from the 10<sup>th</sup> generation transplantation (M-P). Magnification 400x.

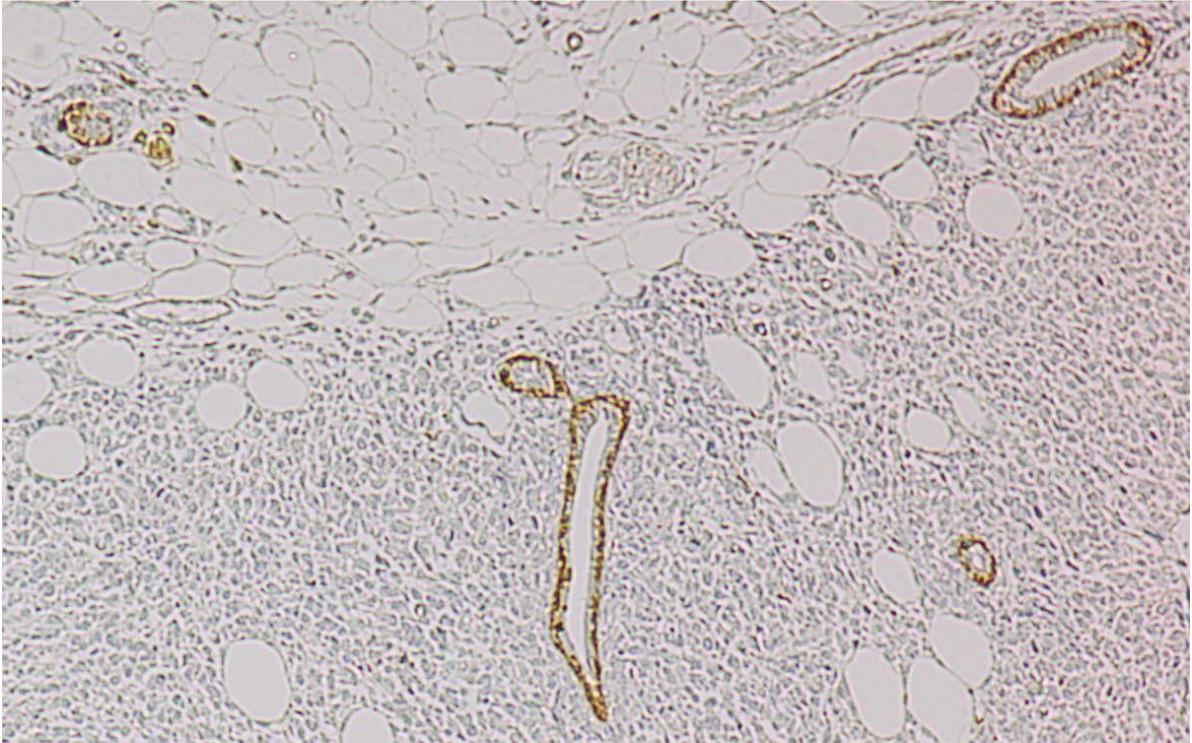


Figure 3: GLI1 induced tumor stained with antibody against K5 showed that the basal cells in remaining normal duct in the tumor tissue were stained with K5. Magnification 40x.

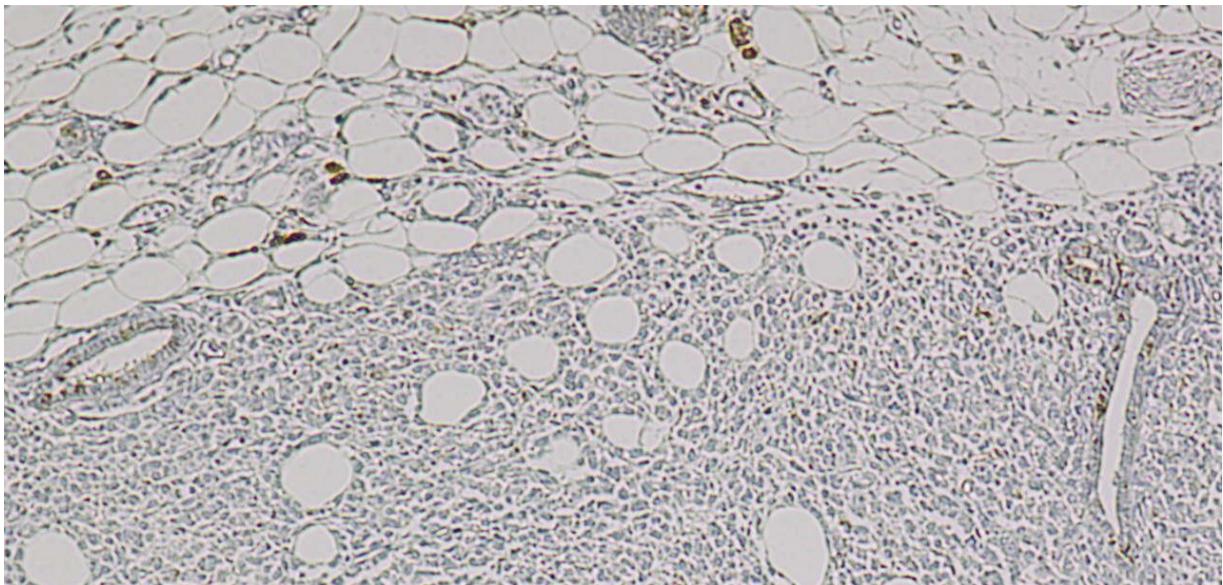


Figure 4: GLI1 induced tumor stained with antibody against K6 showed that the basal cells in remaining normal duct in the tumor tissue were stained with K6. Magnification 40x.

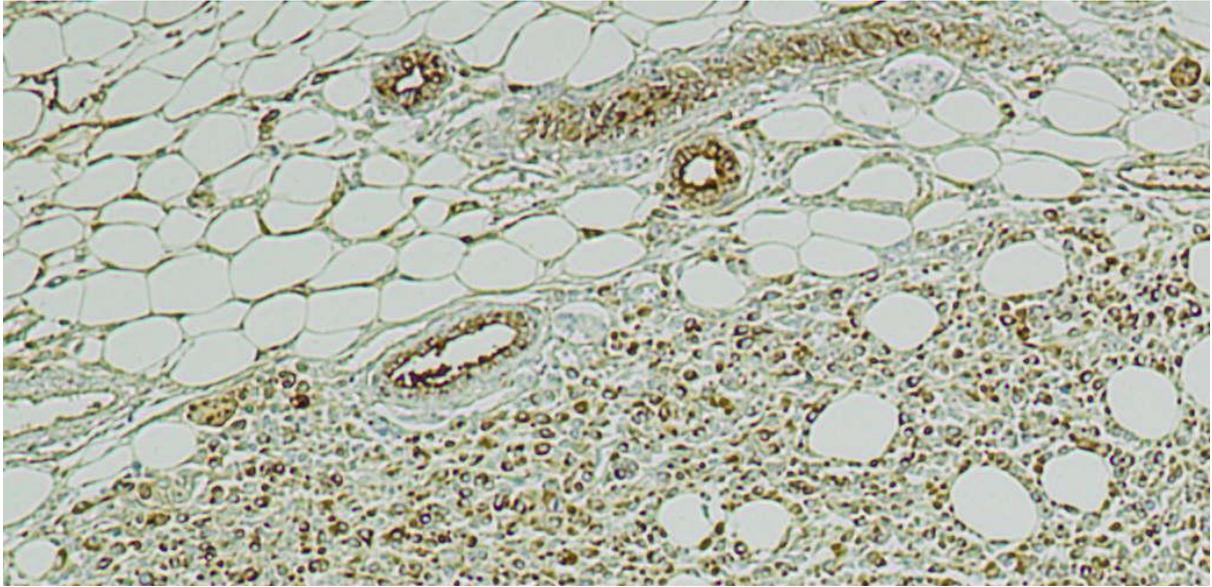


Figure 5: GLI1 induced tumor stained with antibody against K18 showed that the luminal cells in remaining normal duct in the tumor tissue were stained with K18. Magnification 40x.

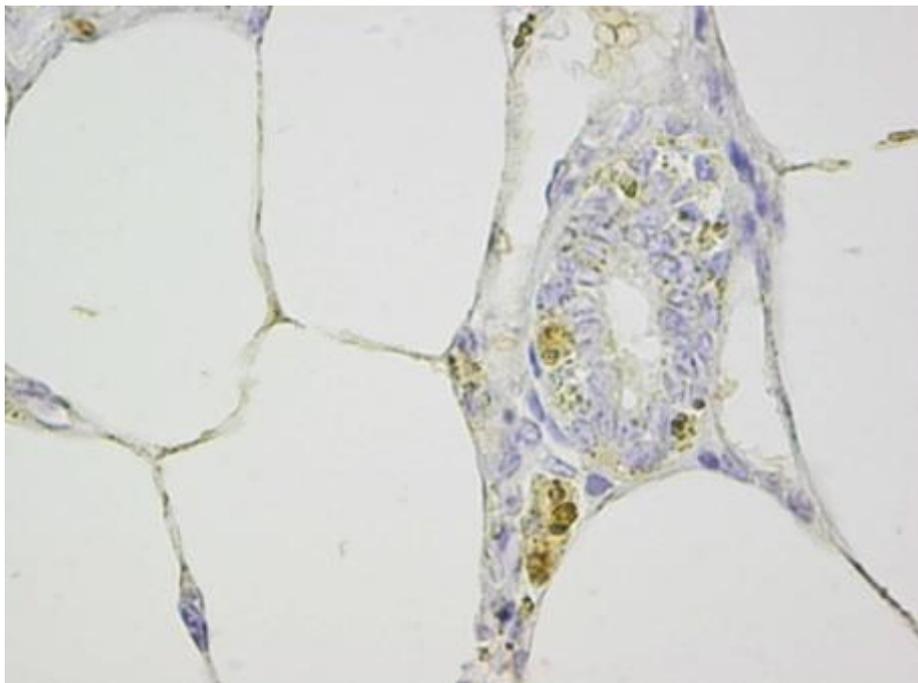


Figure 6: GLI1 induced tumor stained with antibody against Egfr showed positive staining in a normal duct in the tumor tissue was stained with Egfr. Magnification 400x.

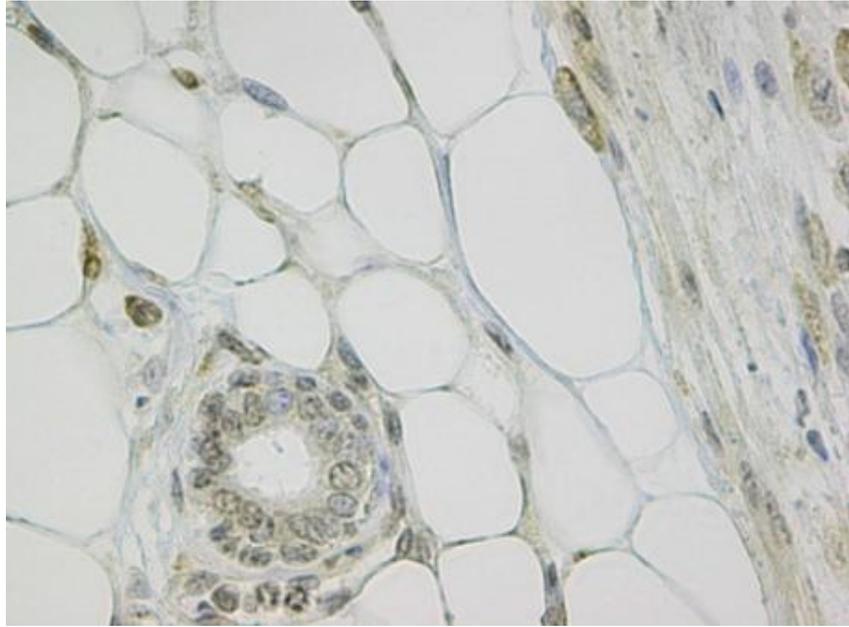


Figure 7: GLI1 induced tumor stained with antibody against Her2 showed positive staining in a normal duct in and in the tumor tissue was stained with K6. Magnification 40x.



Norwegian University  
of Life Sciences

Postboks 5003  
NO-1432 Ås, Norway  
+47 67 23 00 00  
[www.nmbu.no](http://www.nmbu.no)