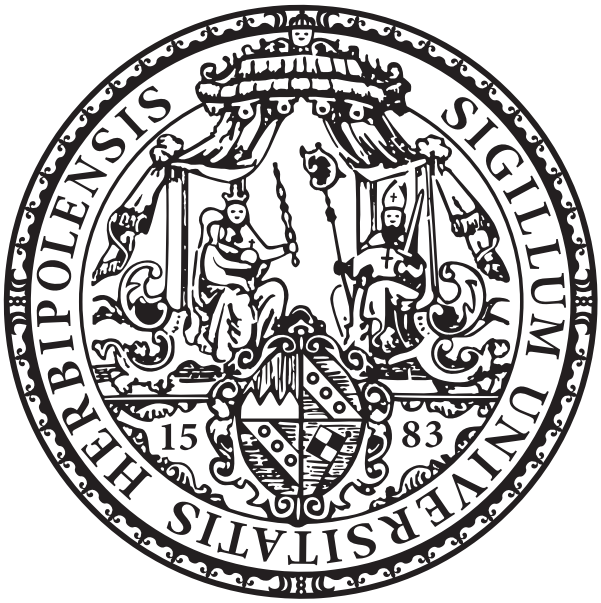
Identification of factors influencing 17beta-estradiol metabolism in female mammary gland



Dissertation

zur Erlangung des naturwissenschaftlichen Doktorgrades der
Julius-Maximilians-Universität Würzburg

vorgelegt von

Katja Schmalbach
aus Cottbus

Würzburg 2014

Eingereicht bei der Fakultät für Chemie und Pharmazie am

_____

Gutachter der schriftlichen Arbeit

1. Gutachter: _____

2. Gutachter: _____

Prüfer des öffentlichen Promotionskolloquiums

1. Prüfer:_____

2. Prüfer:_____

3. Prüfer:_____

Datum des öffentlichen Promotionskolloquiums

_____

Doktorurkunde ausgehändigt am

_____

Danksagung

Ich danke...

... Prof. Dr. Leane Lehmann für die Überlassung des interessanten Themas, die sehr lehrreiche Zeit und die wertvollen Anregungen und Diskussionen, die zum Gelingen dieser Arbeit beigetragen haben.

... Prof. Dr. Dandekar und Dr. Cecil für Rat und Tat zum Netzwerkmodell.

... Prof. Dr. Ickstadt und Claudia Köllman, Helene König, Leo Geppert für Rat und Tat in statistischen Fragen.

... Dr. Jason Bielas for initial training of the Random Mutation Capture Assay.

... Dr. Harald Esch für koordinative Unterstützung und Kontakt zu den Ärzten.

... Prof. Dr. Eckert, Dr. Waldhofen, Dr. Steinmetz für die Ermöglichung der Gewebespenden.

... Allen *IsoCross* Projektmitgliedern für die inspiriereden Diskussionen während den Projektmeetings.

... Anne Scheffler, Benjamin Spielmann, Jan Fuchs für die anregenden Diskussionen rund um den RMCA.

... Annette Albrecht und Daniela Martinez für die Unterstützung in der Zellkultur.

... Susanne Futh für Hilfe bei den Präparationen der Gewebespenden.

... Gesche Fischer und Benjamin Spielmann für ihre engagierte Mitarbeit im Rahmen ihrer wissenschaftlichen Abschlussarbeiten.

... Dem gesamten früheren und jetztigen Arbeitskreis Lehmann für die Diskussionen während Lehrstuhl-internen Seminaren und die gute Zeit auch außerhalb des Labors.

... Dem Deutschen Akademischen Austauschdienst für die finanzielle Unterstützung während meines Aufenthaltes in Labor von Dr. Bielas am Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

... Der Studienstiftung des Deutschen Volkes für die finanzielle Unterstützung meiner Doktorarbeit.

... Sean Kelly and Hans Jung for reading through this.

... Jan für das Korrekturlesen, Zuhören und immer Daseins.

... Ganz besonders meinen Eltern und meiner Oma Lisette, die mich auf meinem gesamten Weg zu dieser Arbeit immer unterstützt haben.

Contents

1 Abbreviations **VI**

2 Introduction **1**

2.1 Morphology, proliferation, differentiation and apoptosis of the human mammary gland . . . 1
2.2 E2 in normal human breast tissue . . . 5
2.2.1 Formation of E1/E2 . . . 5
2.2.2 E1/E2 oxidative metabolism and conjugation . . . 8
2.2.3 Transcription factors . . . 13
2.3 Cellular stress . . . 15
2.3.1 Metabolic networks . . . 16
2.3.2 Determination of spontaneous mutation frequency and mutation spectra 17
2.4 Factors influencing breast cancer risk by interaction with E2 biosynthesis and metabolism . . . 19
2.4.1 Polymorphisms in genes involved in E2 metabolism . . . 20
2.4.1.1 NQO Pro187Ser SNP . . . 21
2.4.1.2 GSTT1 CNP . . . 23
2.4.1.3 COMT Val108/158Met SNP . . . 23

3 Objectives **24**

4 Problemstellung **27**

5 Materials and Methods **29**

5.1 Materials . . . 29
5.1.1 Equipment and laboratory consumables . . . 29
5.1.2 Laboratory consumables . . . 31
5.1.3 Chemicals and enzymes . . . 32
5.1.4 Buffers and solutions . . . 34
5.1.5 Commercial kits and reagents . . . 37
5.1.6 Customized TaqMan® low density arrays . . . 37
5.1.7 Software . . . 41
5.2 Methods . . . 41
5.2.1 Cell culture . . . 41
5.2.2 Preparation of fresh mammary gland tissue . . . 42
5.2.3 Grinding of mammary gland tissue . . . 43

5.2.4 Cryostat sectioning . . . 43
5.2.4.1 Section staining . . . 43
5.2.5 Laser capture microdissection . . . 44
5.2.6 RNA isolation . . . 45
5.2.7 Isolation of total DNA . . . 45
5.2.8 RNA/DNA quantification . . . 46
5.2.9 Reverse transcription . . . 47
5.2.10 Transcript level quantification . . . 47
5.2.10.1 TaqMan® Gene Expression Assays . . . 48
5.2.10.2 Preamplification and pre-preamplification . . . 48
5.2.10.3 TLDA . . . 49
5.2.11 Determination of polymorphisms . . . 50
5.2.11.1 Allelic discrimination . . . 50
5.2.11.2 RFL PCR . . . 51
5.2.11.3 Determination of CNP . . . 52
5.2.12 RMCA . . . 53
5.2.12.1 Synthesis of the uracil containing biotinylated DNA-probe . . . 53
5.2.12.2 DNA restriction digest of genomic DNA . . . 54
5.2.12.3 Microcon purification of DNA restriction Digest . . . 55
5.2.12.4 Hybridization . . . 55
5.2.12.5 *Taq*I digestion . . . 55
5.2.12.6 Copy number quantification – competitive PCR . . . 56
5.2.12.7 Copy number quantification – TaqMan® probe-based real time PCR . . . 58
5.2.12.8 Mutation detection . . . 59
5.2.12.9 Calculation of mutation frequency . . . 60
5.2.13 Agarose gel electrophoresis . . . 60
5.2.13.1 Densitometrical evaluation of agarose gels . . . 61
5.2.13.2 PCR product purification . . . 61
5.2.14 E2 and E1 breast tissue level . . . 61
5.2.15 Statistics . . . 61
5.2.16 Correlation and cluster analysis . . . 63
5.2.17 Principal component analysis and linear regression . . . 64
5.2.18 Calculation of Hardy-Weinberg Equilibrium . . . 65
5.2.19 Metabolic Network . . . 66

6 Results **67**

6.1 Characterization of mammary gland samples 67
6.1.1 Questionnaire 68
6.1.2 Morphology 68
6.1.3 Polymorphisms 70
6.1.3.1 Method establishment and application for the detection of SNPs 71
6.1.3.1.1 RFL PCR COMT Val108/158Met 72
6.1.3.1.2 Allelic discrimination COMT Val108/158Met . . . 73
6.1.3.1.3 Allelic discrimination NQO1 Pro187Ser 75
6.1.3.2 Method establishment and application for the determination of CNP 76
6.1.3.3 Genotypes of human cell lines and human mammary gland tissue samples 77
6.2 Establishment of mRNA level quantification using TaqMan® Low Density Array 81
6.2.1 Selection of transcripts 81
6.2.2 Selection of TLDA format 83
6.2.2.1 Maximum cDNA per slot 83
6.2.2.2 Variations within one slot 84
6.2.3 Variations of one replicate between two slots and two cards 87
6.2.4 Linearity of preamplification and pre-preamplification 92
6.2.4.1 Preamplification 95
6.2.4.2 Pre-preamplification 98
6.2.4.3 Preamplification for TaqMan® Assay 101
6.2.5 Influence of sample-taking sites on transcript levels 102
6.2.5.1 Sample-taking sides 102
6.2.5.2 Sample-taking sites 102
6.2.5.3 Discussion 103
6.3 Transcript levels in mammary gland tissue homogenate 105
6.3.1 Qualitative differences of transcript levels 105
6.3.2 Comparison of transcript levels of mammary gland tissue to MCF-7 cells 106
6.4 Preliminary establishment of transcript level quantification in epithelial cells 114
6.4.1 Laser capture microdissection 114
6.4.2 Quantity of RNA 115
6.4.3 Comparability of transcript levels in two MECs+GECs fractions . . . 117
6.4.4 Transcript levels in epithelial cells and homogenate 119
6.4.5 Comparison of transcript level pattern of MECs+GECs and homogenate 122

6.5 Influence of transcript levels and polymorphisms on genotoxic stress 125
6.5.1 Metabolic network 125
6.5.1.1 Reactions, settings and input data 125
6.5.1.2 Plausibility of the model 127
6.5.1.3 Impact of polymorphisms on respective enzyme fluxes 127
6.5.1.4 Impact of mRNA levels on DNA adduct formation fluxes . . 131
6.5.1.5 Impact of mRNA levels and polymorphisms on DNA adduct formation fluxes 134
6.5.1.6 Limitations and discussion of the metabolic network 135
6.5.1.6.1 Limitations of the metabolic network 136
6.5.1.6.2 Discussion of the metabolic network 137
6.5.2 Random Mutation Capture Assay 138
6.5.2.1 RMCA of TP53 138
6.5.2.2 SMFs of cell lines and normal human mammary gland tissue 142
6.5.2.3 Optimization of the RMCA 143
6.5.2.3.1 Quantification of copy numbers 143
6.5.2.3.2 Mutant detection 144
6.5.2.3.3 SMF in MCF-7 cells 145
6.5.2.3.4 Summary and Discussion 145
6.6 Statistical evaluation of factors interacting with E1/E2 metabolism 146
6.6.1 Correlation analysis of transcripts 146
6.6.2 Influence of individual factors on transcript levels 150
6.6.2.1 Associations of lobule type and age 151
6.6.2.2 Associations of BMI and smoking habits with transcript levels 153
6.6.3 Linear regression models 154
6.6.3.1 Factors influencing E1 and E2 tissue level 155
6.6.3.2 Factors influencing DNA adduct formation 158

7 Summary **161**

8 Zusammenfassung **167**

9 Appendix **195**
9.1 Questionnaire 195
9.2 Lobule types and characterization 196
9.3 TLDA 206
9.3.1 Preamplification 211
9.3.2 Sample-taking sites 219

9.3.3 Calculation of expected ΔCt values 221
9.4 Transcript levels of homogenate 221
9.5 Transcript level of epithelial cells 227
9.6 Metabolic Network 230
9.7 Statistic mRNA level breast tissue samples 232

1 Abbreviations

AHR	Aryl hydrocarbon receptor
approx.	Approximately
ARE	Antioxidant responsive element
AREG	Amphiregulin
ARNT	Aryl hydrocarbon receptor nuclear translocator
BAD	BCL2 antagonist of cell death
BAX	BCL2-associated X protein
BMF	Proapoptotic BH3-only protein
BMI	Body mass index
BSA	Bovine serum albumin
CCND1	Cyclin D1
CDKN1A	Cyclin-dependent kinase inhibitors 1A
CDKN1B	Cyclin-dependent kinase inhibitors 1B
CNP	Copy number polymorphism
COMT	Catechol-*O*-methyltransferase
Ct	Cycle threshold
CYP	Cytochrome P450 monooxygenase
CYP19A1	Aromatase
DHEA	Dihydroepiandrosterone
DHEAS	Dihydroepiandrosterone sulfate
DNA	Deoxyribonucleic acid
E2	17β estradiol
E1	Estrone
e.g.	Exempli gratia
EGFR	Epidermal growth factor receptor
ESR1	Estrogen receptor α
ESR2	Estrogen receptor β
ERE	Estrogen responsive element
FCS	Fetal calf serum
GADD45A	Growth arrest and DNA-damage-inducible protein 45 alpha
GATA3	GATA binding protein 3
GCLC	γ-glutamyl-cysteine ligase
GEC	Glandular epithelial cell
GSH	Glutathione

GST	Glutathione-S-transferase
H&E	Hematoxylin and eosin stain
HSD	Hydroxysteroid dehydrogenase
HWE	Hardy-Weinberg Equilibrium
i.e.	Id est
IGF	Insulin-like growth factor
IGFR	Insulin-like growth factor receptor
KEGG	Kyoto Encyclopedia of Genes and Genomes
Leu	Leucine
MEC	Myoepithelial cell
Met	Methionine
Min	Minute(s)
MKI67	Proliferation-related Ki-67 antigen
n0	Initial copy number
NADPH	Nicotinamide adenine dinucleotide phosphate
Ncf-water	nuclease- and nucleic acid-free water
NFE2L2	Nuclear factor (erythroid-derived 2)-like 2
NTCP	Sodium/taurocholate cotransporting polypeptide
NQO1	NADPH-quinone oxidoreductase
NR1I2	Nuclear receptor-sub family 1-group I-member 2
NR1I3	Nuclear receptor-sub family 1-group I-member 3
OAT	Organic anion transporters
OATP	Organic anion-transporting polypeptide
PAPS	3'-Phosphoadenosine-5'-phosphosulfate
PCR	Polymerase chain reaction
PGR	Progesterone receptor
Pro	Proline
R^2	Coefficient of correlation
RMCA	Random Mutation Capture Assay
RNA	Ribonucleic acid
mRNA	messenger RNA
SAM	S-adenosylmethionine
SD	Standard deviation
Sec	Seconds
Ser	Serine
SFN	14-3-3σ

SLC	Solute carrier
SMF	Spontaneous mutation frequency
SNP	Single nucleotide polymorphism
SOAT	Sodium-dependent organic anion transporter
STS	Steroid sulfatase
SULT	Sulfotransferase
TGFB1	Transforming growth factor β
TFF1	Trefoil factor 1
TLDA	TaqMan® Low Density Array
TP53	Tumor suppressor protein 53
UDG	Uracil-DNA glycosylase
UDPGA	Uridine diphosphate glucuronic acid
UGT	Uridine 5'-diphospho-glucuronosyltransferase
Val	Valine
vs	Versus
WNT	wingless-related MMTV integration site

2 Introduction

In 2010, 70,340 women developed breast cancer in Germany and the prognosis for 2014 is about 75,000 new cases (RKI and GEKID, 2013). At diagnosis every fourth woman is younger than 55 and every tenth woman is younger than 45 (RKI and GEKID, 2013). Only a small share of breast cancer patients has heritable mutations and therefore an increased risk to develop breast cancer (RKI and GEKID, 2013), while the rest of them may acquire genetic mutations during their lifetime.

A contributing factor to breast cancer development is the female sex hormone 17β-estradiol (E2), produced naturally in the body, because some of the well-established risk factors for breast cancer, e.g. early onset of menarche (<12 years) and late menopause (>55 years) are related to lifetime exposure of mammary gland tissue to E2.

E2 acts in two different ways (summarized in Yager, 2014; Santen et al., 2014): (i) E2 can be activated to reactive metabolites by enzymes of the cytochrome P450 family present in mammary gland tissue. These reactive metabolites are known to damage DNA. (ii) The stimulation of the estrogen receptor α (ESR1) by E2 increases cell proliferation which can also increase the mutation frequency and may contribute to clonal expansion of damaged cells.

Besides early onset of menarche and late menopause, there are other known risk factors for breast cancer development (Chapter 2.4) which may also influence E2 metabolism in mammary gland tissue (Chapter 2.2). In order to investigate those factors the morphology of the mammary gland tissue (Chapter 2.1), as well as the E2 metabolism in mammary gland tissue leading to cellular stress (Chapter 2.3) must be taken into account.

2.1 Morphology, proliferation, differentiation and apoptosis of the human mammary gland

The adult mammary gland tissue is characterized by a network of ducts ending in small clusters of terminal ductal lobular units (Figure 1). Ducts and lobules are surrounded by fibrous stroma and fat cells. The intralobular stromal cells are located between the ductules within a lobule and the interlobular stroma is located between lobules (summarized in Bocker et al., 2009).

The stage of differentiation of ducts and lobules in the adult breast is dependent on age and parity. In puberty the ducts grow, divide, and form club-shaped terminal end buds (Figure 2). Terminal end buds give origin to new branches, twigs and small ductules or alveolar buds. Lobule formation in the female breast occurs within 12 years after onset of the first menstrual period (Russo and Russo, 2004). Alveolar buds cluster around a terminal duct,

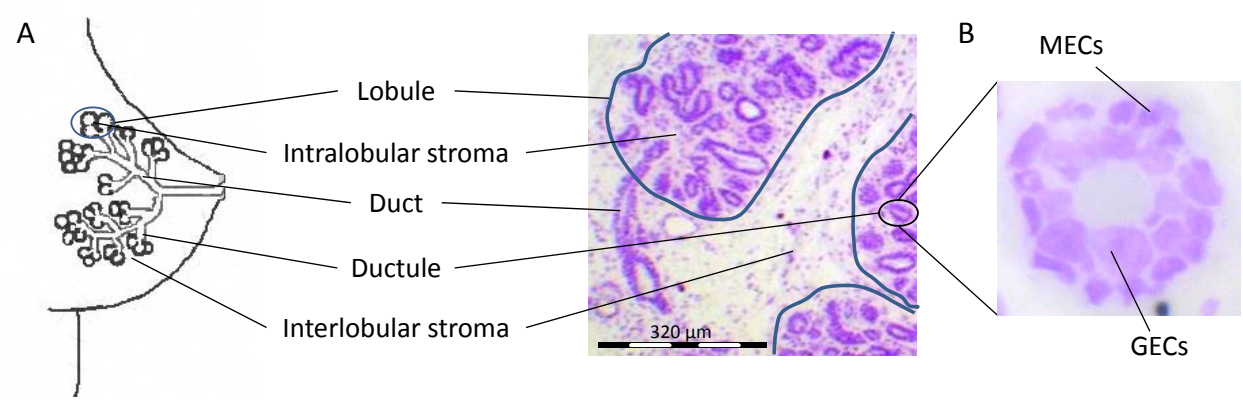


Figure 1: Morphology of the female breast (modified according to Bocker et al., 2009). (A) The intralobular stromal cells are located between the ductules within a lobule. The interlobular stroma is located between lobules. (B) Lobules contain ductules, each lined with glandular epithelial cells (GECs) encircled by myoepithelial cells (MECs). MECs are located at the basal membrane.

forming the lobule type 1 (Figure 2) and each cluster is composed of approx. 11 alveolar buds (Russo and Russo, 2004). The breast of nulliparous women is characterized predominately by lobule type 1 (Russo and Russo, 2004), although occasionally, lobules with higher grade of differentiation are detected (lobule type 2 and 3). The transition from lobule type 1 to type 2, and from type 2 to type 3, is a gradual process of sprouting of new alveolar buds (called ductules). The number of ductules per lobule is increased from 11±6 in lobule type 1 to 47±12 and 81±17 in lobule types 2 and 3, respectively (Russo and Russo, 2004). The increase in number is accompanied by an increase in size of the lobules and a reduction in size of each ductule (Russo and Russo, 2004). In breast tissue of parous women the lobule types 2 and 3 are predominant during reproductive years, but decrease back to lobule type 1 after fourth decade of life. Moreover, the age-dependent involution of mammary gland tissue of parous women has been described to pre-date the onset menopause (Hutson et al., 1985).

Thus, during lifetime the human mammary gland tissue undergoes several processes of proliferation, differentiation and involution.

Proliferation

The glandular epithelium is the main site of cell proliferation (summarized in Clarke, 2006) and stimulation of ESR1 by E2 increases cell proliferation. Only 20% of the cells in the glandular epithelium proliferate and proliferating glandular epithelial cells (GECs) do not contain ESR1 but they are in close proximity to ESR1 and progesterone receptor (PGR) positive GECs (summarized in Clarke, 2006). Only 10-20% of GECs express ESR1 whereas ESRβ (ESR2) is detected in the majority of GECs, myoepithelial, and stromal cells (summarized in Clarke, 2006). Additionally, ESR1, ESR2 and PGR are differentially expressed in nulliparous vs. parous women (Asztalos et al., 2010). According to current knowledge, proliferation of GECs is controlled by paracrine signals from both epithelial and

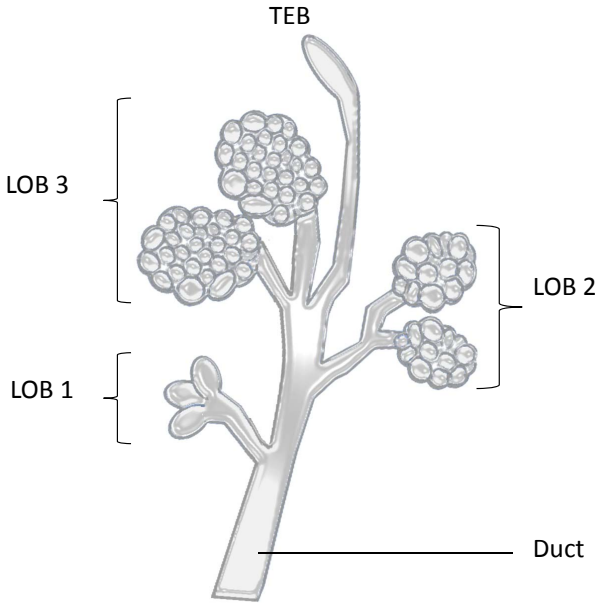


Figure 2: Scheme of lobular structure of the human mammary gland. TEB, terminal end bud. LOB, lobule type.

stromal cells (summarized in Visvader and Stingl, 2014).

E2 is the most biologically active estrogen (summarized in Samavat and Kurzer, 2014), but also E1 (60% relative binding affinity of E2 to ESR1, Kuiper et al., 1997), metabolites of E2 (2-hydroxy(OH)-E2 and 4-OH-E2: 7% and 13% relative binding affinity of E2 to ESR1, Kuiper et al., 1997) and precursors in steroidogenesis, such as dihydroepiandrosterone (DHEA) possess estrogenic activity (0.04% relative binding affinity of E2 to ESR1, Kuiper et al., 1997). Ligand-activated ESR1 induces e.g. trefoil factor 1 (TFF1, summarized in Ribieras et al., 1998), PGR (summarized in Clarke, 2006), cyclin D1 (CCND1, Eigeliene et al., 2008b) and amphiregulin (AREG, McBryan et al., 2008). TFF1, PGR and CCND1 contain full estrogen responsive elements (EREs) or half-ERE sequences in proximity to Sp1-binding sites (summarized in O'Lone et al., 2004). A direct binding of ligand-activated ESR1 to DNA sequence of AREG is assumed (McBryan et al., 2008). AREG transmits the signal to the stroma by activating stromal tyrosine kinase epidermal growth factor receptor (EGFR)/ErbB1 pathway (Figure 3). In response, activated stromal cells release growth factors, such as insulin-like growth factor (IGF1), inducing cell proliferation in adjacent ESR1 negative epithelial cells, expressing IGF receptor (IGFR, Figure 3). This process is counteracted by the transforming growth factor β (TGFB1) which is secreted by GECs to inhibit branching and ductal growth via stromal TGFB receptor (summarized in Lanigan et al., 2007).

The cyclin-dependent kinase inhibitors 1A (CDKN1A) and 1B (CDKN1B) encoding the cyclin-dependent kinase p21 and p27, respectively, are also involved in proliferation. p21 neg-

atively modulates cell cycle progression (summarized in Gartel and Radhakrishnan, 2005), while an induction of CDKN1B causes G1 cell cycle arrest (summarized in Burhans and Heintz, 2009). Additionally, transcript level of proliferation-related Ki-67 antigen (MKI67) is a marker for proliferating epithelial cells of the mammary gland (Eigeliene et al., 2008a, Figure 3).

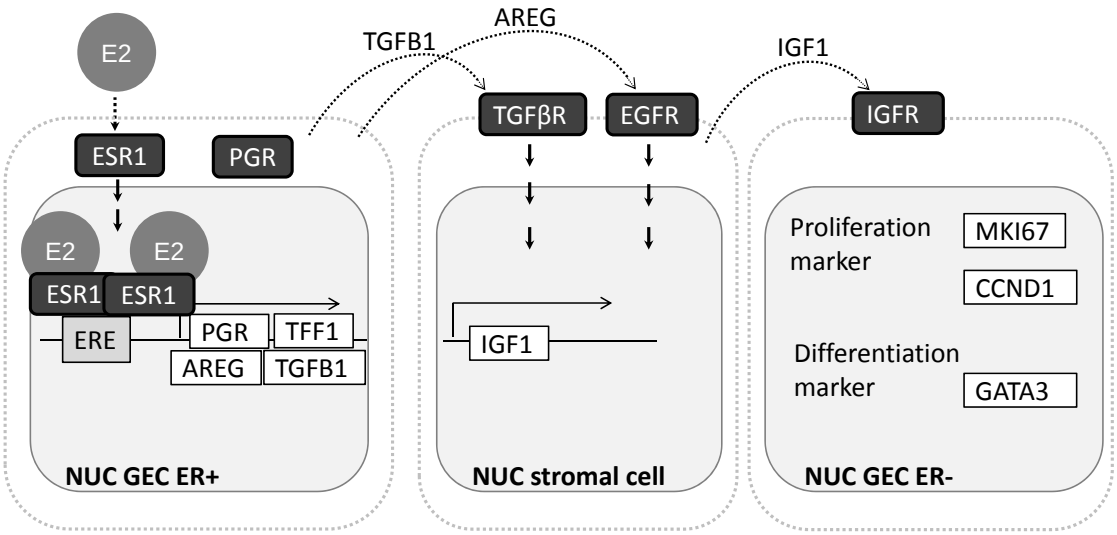


Figure 3: Simplified scheme of ESR1 activation and proliferation in human mammary gland. NUC, nucleus. ER+, ESR1 containing cells. ER-, cells without ESR1. TGFβR, TGFB receptor.

Differentiation

A key inducer of GEC differentiation seems to be the GATA binding protein 3 (GATA3, Figure 3). Higher levels of GATA3 may drive cells into a more differentiated state (summarized in Visvader, 2009). Further mediators involved in differentiation and side branching are wingless-related MMTV integration site-(WNT)4 and 5A. WNT4 is a downstream effector of PGR required for the ductal side-branching in the breast (summarized in Kim et al., 2009). PGR is co-localized in GECs expressing ESR1 (summarized in Visvader and Stingl, 2014). The expression of WNT5A is directly regulated by TGFB1 and mediates inhibiting effects of TGFB1 on ductal side extension (summarized in Serra et al., 2011). Thus, both WNTs have opposing effects on ductal side branching and extension during mammary gland development.

Involution

The lobular involution occurs postlactational, which will be not addressed in this work, and age-related: The age-related process involves replacement of epithelial cells and interlobular stromal tissue by adipose tissue (summarized in Macias and Hinck, 2012). Therefore, it is important for characterization of the lobule type in normal mammary gland tissue. The proapoptotic BH3-only protein (BMF), as well as the apoptosis BCL2 antagonist of cell death

(BAD) and the BCL2-associated X protein (BAX) are potential mediators of mammary gland involution (Schmelzle et al., 2007; Schorr et al., 1999). BMF is also targeted by TGFB1 dependent signals (summarized in Pinon et al., 2008). BMF transcript levels are up regulated during anoikis and acinar morphogenesis of mammary epithelial cells (Schmelzle et al., 2007). The tumor suppressor protein 53 (TP53) induces cell cycle arrest or apoptosis depending on activated target genes (induction of p21 or transcriptional activation of BCL2 family members, respectively, summarized in Insinga et al., 2014). CDKN1A is a downstream target of TP53 and both accumulate in senescent cells (summarized in Burhans and Heintz, 2009). However, immunohistochemical staining revealed no positive signals for TP53 neither in stromal nor in epithelial cells of normal mammary gland (Mao et al., 2010b; Milicevic et al., 2014).

2.2 E2 in normal human breast tissue

E2 is the most biological active estrogen, primarily secreted in granulosa cells of the ovaries in premenopausal women (summarized in Samavat and Kurzer, 2014, Chapter 2.2.1), which is then released into blood circulation. Normal human breast tissue, as well as adipose tissue, act as an *intracrine organ with local estrogen synthesis* (summarized in Yaghjyan and Colditz, 2011, Chapter 2.2.1). Furthermore, E2 can be oxidatively metabolized in breast tissue (Chapter 2.2.2). Transcription of E2 metabolizing genes is controlled by transcription factors, which may be also affected by factors influencing breast cancer risk (Chapter 2.2.3).

2.2.1 Formation of E1/E2

The cortex of the adrenal gland is the main source of DHEA and androstenedione (summarized in Blair, 2010), precursors of estrogens. Estrone (E1), E1 sulfate and E2 are synthesized from cholesterol via pregnenolone, 17-hydroxypregnenolone, DHEA and androstenedione in granulosa and theca cells of the ovaries (summarized in Miller and Auchus, 2011). DHEA, DHEA sulfate (DHEAS) and androstenedione are synthesized in andrenal glands (summarized in Miller and Auchus, 2011, Figure 5). Furthermore, estrogens are synthesized in intracrine tissue such as mammary gland and adipose tissue (Simpson, 2003, Figures 5 and 4). DHEAS and E1 sulfate are the dominant sex steroids in blood circulation (summarized in McNamara and Sasano, 2014, Figure 5). E2 is present in blood unbound or non-covalently bound to steroid binding proteins, such as sex hormone binding globulin or albumin: In blood 1.8% of E2 are free, 37.3% are bound to sex hormone binding globulin and 60.8% bound to albumin (Dunn et al., 1981).

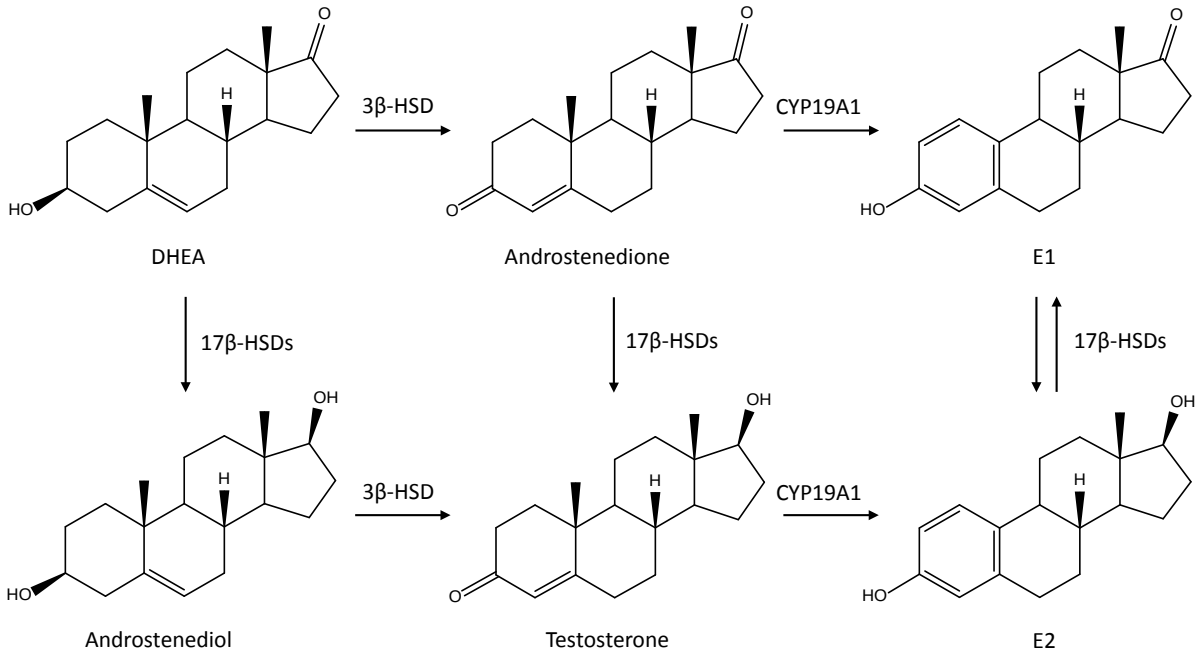


Figure 4: Formation of E1 and E2 (modified according to Blair, 2010). DHEA and androstenediol are metabolized to androstenedione and testosterone, respectively, catalyzed by 3β-hydroxysteroid dehydrogenase (HSD), which are then metabolized (CYP19A1, aromatase) to E1 and E2, respectively. E1 and E2 are converted into each other, catalyzed by 17β-HSDs.

Transport into mammary gland tissue

Free E1 and E2 readily diffuse through plasma membranes, while the negatively charged sulfates and glucuronides need transport e.g. by organic anion-transporting polypeptides (OATPs, summarized in Obaidat et al., 2012) or organic anion transporters (OATs, summarized in Koepsell, 2013). OATs and OATPs are encoded by the genes of the two solute carrier (SLC) gene superfamilies, SLC22A and SLC0, respectively (summarized in Roth et al., 2012).

OATPs are capable of bidirectional transport of the E1/E2 metabolites and precursors. The mechanism of the OATP transport is still unclear, however, the transport is ATP- and sodium-independent. Eight of the eleven isoforms are known to transport E1/E2 metabolites or DHEAS (Table 1). Except for two isoforms (OATP1B1 and OATP1C1), all are expressed in normal or cancer breast tissue, however, most studies investigating normal mammary gland tissue were carried out with normal tissue adjacent to tumor tissue (summarized in Obaidat et al., 2012; Kindla et al., 2011).

The most investigated transporters of the OAT family (OAT1 and OAT3) are known to transport organic anions against a negative membrane potential in exchange for the counter ion α-ketoglutarate. In contrast, OAT7 exhibits a different exchange mechanism using short chain fatty acids such as butyrate as counter ions (summarized in Roth et al., 2012).

Table 1: Known substrates among E1/E2 metabolites and precursors of the OATP transporter family (summarized in Roth et al., 2012), OAT transporter (summarized in Koepsell, 2013) and SLC10 transporter (summarized in Claro da Silva et al., 2013) in normal and cancer breast tissue. E1-3-S, E1-3-sulfate. E2-17β-G, E2-17β-glucuronide. Preg-S, pregnenolone sulfate. N, normal breast tissue. C, breast cancer tissue. n.d., not determined. ?, not known.

Transporter	Gene	Substrate	Expression in mammary gland tissue	
OATP1A2	SLC01A2	E1-3-S, E2-17β-G, DHEAS	C	(summarized in Obaidat et al., 2012)
OATP1B1	SLC01B1	E1-3-S, E2-17β-G, DHEAS	n.d.	(Kindla et al., 2011)
OATP1B3	SLC01B3	E1-3-S, E2-17β-G, DHEAS	N + C	(summarized in Obaidat et al., 2012)
OATP1C1	SLC01C1	E1-3-S, E2-17β-G	n.d.	(Kindla et al., 2011)
OATP2B1	SLC02B1	E1-3-S, DHEAS	N + C	(summarized in Obaidat et al., 2012)
OATP3A1_v1	SLC03A1	E1-3-S	C	(summarized in Obaidat et al., 2012)
OATP4A1	SLC04A1	E1-3-S, E2-17β-G	N + C	(summarized in Obaidat et al., 2012)
OATP4C1	SLC04C1	E1-3-S	N + C	(summarized in Obaidat et al., 2012)
OAT2	SLC22A7	E1-3-S, DHEAS	?	
OAT3	SLC22A8	E1-3-S, E2-17β-G, DHEAS	N	(summarized in Koepsell, 2013)
OAT4	SLC22A11	E1-3-S, DHEAS	?	
OAT7	SLC22A9	E1-3-S, DHEAS	?	
NTCP	SLC10A1	E1-3-S, DHEAS	?	
SOAT	SLC10A6	E1-3-S, DHEAS, Preg-S	N	(summarized in Claro da Silva et al., 2013)

Four of the nine isoforms are known to transport E1/E2 metabolites or DHEAS, but only OAT3 is expressed in normal breast tissue (summarized in Koepsell, 2013). No information was available on the expression levels in mammary gland tissue for the other isoforms (Table 1).

Furthermore, E1 sulfate and DHEAS are transported by two transporters of the SLC10 family: sodium/taurocholate cotransporting polypeptide (NTCP) and sodium-dependent organic anion transporter (SOAT, summarized in Claro da Silva et al., 2013). Only SOAT is known to be expressed in normal breast tissue, NTCP is present in the liver (summarized in Claro da Silva et al., 2013).

Synthesis of E2 in mammary gland tissue

In breast tissue E1/E2 is directly synthesized from androstenedione/testosterone, catalyzed by aromatase (CYP19A1, summarized in Blair, 2010, Figure 5). Additionally, in mammary gland, DHEA or DHEAS, after de-sulfonation catalyzed by steroid sulfatase (STS), are metabolized to androstenedione, catalyzed by 3β-hydroxysteroid dehydrogenase (HSD), which is then further metabolized to E1. E1 and E2 are converted into each other: E1 is converted to E2 in breast tissue by 17β-HSD(HSD17B)-1, 5, 7, 12 and HSD17B2 and 14 converts E2 back to E1 (summarized in Blair, 2010). HSD17B1, 2, 5, 7, 12 and 14 are expressed in epithelial cells of acini and/or ducts as well as in the stromal cells in the breast (summarized in Blair, 2010; Sivik et al., 2012; Li et al., 2009).

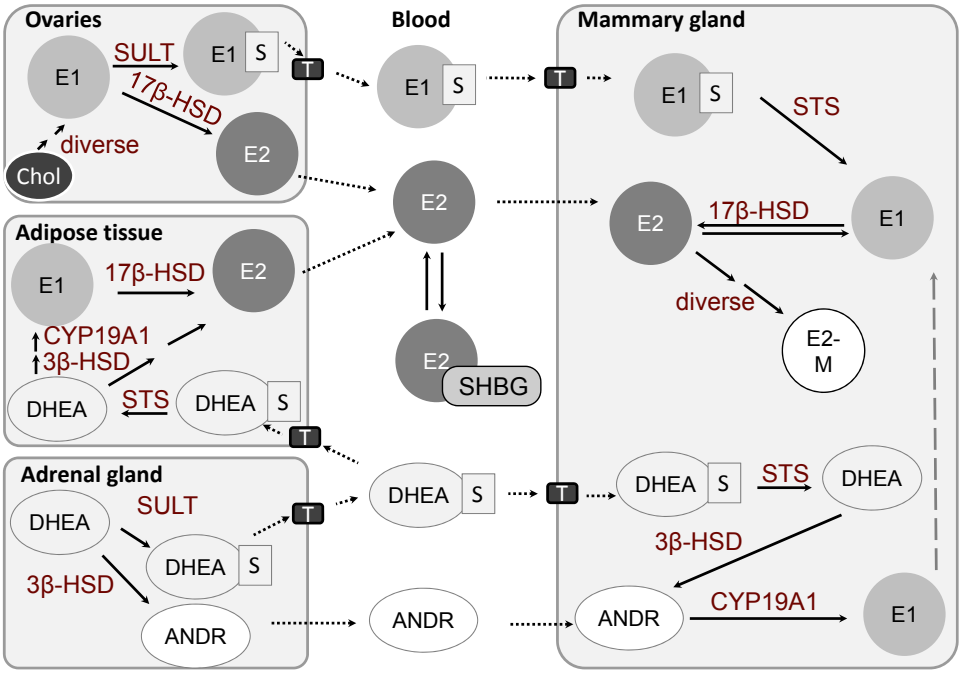


Figure 5: Formation E1 and E2 in premenopausal women. T, transporter. M, metabolites. S, sulfate. Chol, cholesterol. ANDR, androstenedione. SHBG, sex hormone binding globulin or albumin.

After menopause the serum/plasma levels of E1 sulfate and E2 decrease (Pasqualini et al., 1996) due to ceasing secretion by the ovaries, while the DHEA and DHEAS serum levels (produced in the adrenal glands) are almost unchanged after menopause (summarized in Labrie et al., 2001). Hence, the formation of E1 and E2 by peripheral tissue, such as breast and adipose tissue becomes the most important source of E1 and E2 (summarized in Blair, 2010).

2.2.2 E1/E2 oxidative metabolism and conjugation

E1 and E2 are eliminated from the human body by a variety of metabolic conversions to hormonally inactive or less active metabolites. The glucuronidation and sulfonation of E1 and E2 leads to inactivation. Sulfates serve as a depot of E1 and E2 in the body (Suzuki et al., 2003).

However, some of the hormonally inactive or less active metabolites are reactive against cellular structures and thus are potentially genotoxic:

In human mammary gland tissue E1 and E2 are hydroxylated by cytochrome P450 monooxygenases (CYPs), generating catechol estrogens 2- and 4-HO-E1 and -E2 (summarized in Blair, 2010, Figure 6). The genotoxic catechol estrogens are inactivated by several conjugation reactions, such as glucuronidation, sulfonation and *O*-methylation. These catechol estrogens can be further oxidized to E1(E2)-2,3-quinones and E1(E2)-3,4-quinones,

which may form adducts with cellular structures such as DNA and therefore cause DNA damage (summarized in Cavalieri and Rogan, 2011). The estrogen quinones can be detoxified by conjugation with glutathione (GSH), spontaneously or catalyzed by glutathione-S-transferases (GSTs), or they are reduced to catechol estrogens, catalyzed by NADPH-quinone oxidoreductase (NQO1, Albin et al., 1993; Dawling et al., 2004).

Hydroxylation

CYP1A1, 1A2, 3A4 metabolize E1 and E2 either to 2-HO-E1/E2 or 4-HO-E1/E2. CYP1B1 hydroxylates E1 and E2 at position 4. CYP3A5, 3A7 and CYP2C8 hydroxylate in position 16 (summarized in Blair, 2010). 2-HO-E1(E2) and 4-HO-E1(E2) can be further metabolized to reactive quinones also catalyzed by CYP-isoenzymes or peroxidases (Figure 6).

Only CYP1A1 and CYP1B1 (Lehmann and Wagner, 2008), CYP3A5 (Bandala et al., 2012) are expressed in normal mammary gland tissue. The mRNA (*in situ* hybridization) and protein (immunohistochemistry) of CYP1B1 were detected exclusively in epithelial cells (Muskhelishvili et al., 2001). cDNA expression array data in MCF-10F suggest the presence of CYP3A7 transcript in cultured breast epithelial cells, but no information is available on normal mammary gland tissue (Calaf and Roy, 2007).

mRNA of CYP2C8 was detected in ten out of ten breast cancer samples using reverse-transcriptase polymerase chain reaction (PCR, Knupfer et al., 2004), while no mRNA was detected in 29 out of 29 breast cancer samples and adjacent normal tissue also using reverse-transcriptase PCR (Modugno et al., 2003).

mRNA of CYP1A2 and 3A4 has not been detected in normal mammary gland tissue (Lehmann and Wagner, 2008). Immunohistochemical staining of normal mammary gland tissue with antibodies against CYP3A4, however, revealed positive reactions mostly in epithelial cells (Floriano-Sanchez et al., 2014). Taken together, the data indicate CYP1A1 and 1B1 as the most important CYP isoenzymes for hydroxylation at positions 2 and 4 in normal mammary gland tissue.

Sulfonation

E1 and E2 catechols are mono-sulfonated at position 3 or 17 by sulfotransferases (SULTs), whereby the formation of E2-17-sulfate, the minor metabolite, is only mediated by SULT2A1 (summarized in Blair, 2010). E1 is sulfonated at position 3 (summarized in Blair, 2010). SULT1A1, SULT1A2, SULT1A3, SULT1A4, SULT1E1, SULT2A1 and SULT2B1 are known to sulfonate E1, E2 or their respective metabolites (Adjei and Weinshilboum, 2002). SULT1E1 and SULT2A1 are know to sulfonate E1 and E2, where SULT1E1 has an approx. 100 times lower K_m value compared to SULT2A1 (Adjei and Weinshilboum, 2002).

Table 2: Enzymes and isoenzymes metabolizing E1/E2 or metabolites, their substrates in E1/E2 metabolism and whether the enzymes are expressed in normal mammary gland tissue (+) or not (-). If the localization in the mammary gland tissue is known, the "+" was replaced. T, testosterone. MeO-, methoxy-. COMT, catechol-*O*-methyltransferase. ?, no data available.

Enzyme/ Isoenzymes	Substrate	Mammary gland tissue expression (cell type)	References
CYP19A1	T	epithelium and stroma	Cavalieri et al. 2006; Blair 2010; Li et al. 2009; Sivik et al. 2012
HSD17B1	E1	epithelium and stroma	
HSD17B2	E2	epithelium and stroma	
HSD17B5	E1	epithelium and stroma	
HSD17B7	E1	epithelium and stroma	
HSD17B12	E1	epithelium and stroma	
HSD17B14	E2	epithelium and stroma	
CYP1A1	E1, E2	+	Cavalieri et al. 2006; Lehmann and Wagner 2008; Hachey et al. 2003; Muskhelishvili et al. 2001; Bandala et al. 2012; Knupfer et al. 2004; Modugno et al. 2003
CYP1A2	E1, E2	-	
CYP1B1	E1, E2	epithelium	
CYP3A4	E1, E2	-	
CYP1A5	E1, E2	+	
CYP3A7	E1, E2	?	
CYP2C8	E1, E2	?	
SULT1A1	E2, 2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2	+	Adjei and Weinshilboum 2002; Aust et al. 2005; Li et al. 2009; Lehmann and Wagner 2008; Dumas et al. 2008
SULT1A2	E2, 2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2	+	
SULT1A3/4	2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2	+	
SULT1E1	E1, E2, 2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2	epithelium and stroma	
SULT2A1	E1, E2, 2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2	-	
SULT2B1	E2, 2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2	+	
STS	E1-S, E2-S	epithelium and stroma	Blair 2010; Li et al. 2009
UGT1A1	E1, E2, 2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2, 2-MeO-E1, 4-MeO-E1, 2-MeO-E2, 4-MeO-E2	-	Lepine et al. 2004; Chouinard et al. 2006; Starlard-Davenport et al. 2008; Lehmann and Wagner 2008; Thibaudeau et al. 2006
UGT1A3	E1, E2, 2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2, 2-MeO-E1, 4-MeO-E1, 2-MeO-E2, 4-MeO-E2	+	
UGT1A8	E1, E2, 2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2, 2-MeO-E1, 4-MeO-E1, 2-MeO-E2, 4-MeO-E2	epithelium	
UGT1A9	E1, 2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2, 2-MeO-E1, 4-MeO-E1, 2-MeO-E2, 4-MeO-E2	epithelium	
UGT1A10	E1, E2, 4-HO-E2, 2-MeO-E1, 4-MeO-E1, 2-MeO-E2, 4-MeO-E2	+	
UGT2B7	E2, 2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2	epithelium	
COMT	2-HO-E1, 4-HO-E1, 2-HO-E2, 4-HO-E2	epithelium	Lehmann and Wagner 2008; Weisz et al. 2000
NQO1	E1/E2-quinones	epithelium	Lehmann and Wagner 2008; Siegel and Ross 2000
GSTA	?	-	Cairns et al. 1992; Oguztuzun et al. 2011; Hachey et al. 2003; Terrier et al. 1990
GSTM1	?	epithelium	
GSTT1	?	+	
GSTP1	E1/E2-quinones	epithelium	

SULT1A1 and SULT1A2 only sulfonate E2 (Adjei and Weinshilboum, 2002). All of these four isoenzymes and SULT1A3 and SULT2B1 sulfonate E1/E2 catechols, while SULT2B1 has the highest affinity (Adjei and Weinshilboum, 2002). However, SULT2B1 more specifically sulfates DHEA (Meloche and Falany, 2001) and does not sulfonate E1 or E2 (Adjei and Weinshilboum, 2002). SULT1A1, SULT1A2, SULT1A3, SULT1A4 and SULT1E1 are expressed in normal mammary gland tissue (Lehmann and Wagner, 2008; Li et al., 2009; Aust et al., 2005). The transcript of SULT2A1 was not detected in normal mammary gland RNA (Lehmann and Wagner, 2008), but the result was only based on one sample. SULT2B1 was detected in normal breast tissue adjacent to tumor tissue (Dumas et al., 2008).

All E1/E2-sulfates are de-sulfonated by STS, which is expressed in epithelial cells and stroma of normal mammary gland tissue (Li et al., 2009).

Glucuronidation

Members of the uridine 5'-diphospho-glucuronosyltransferase (UGT) family, particularly UGT1A1, UGT1A3, UGT1A8, UGT1A9, UGT1A10 and UGT2B7, are known to catalyze the glucuronidation of E1 (at position 3), E2 (at position 3 and 17) and the respective metabolites (Lepine et al., 2004). UGT1A1 is known to catalyze the glucuronidation of E2 at position 3 and UGT1A3 and 2B7 at position 17 (summarized in Blair, 2010). UGT2B7 only catalyzes the glucuronidation of E2 and the E1/E2 catechols, but not the glucuronidation of the methylated catechols and E1 (Lepine et al., 2004). In contrast, UGT1A9 does not catalyze the glucuronidation of E2, but catalyzes the glucuronidation of all other E1/E2 metabolites and E1 (Lepine et al., 2004). The glucuronidation of 2-HO-E1(E2) and methoxy estrogens occurs predominately at position 3 and of 4-HO-E1(E2) at position 4 (summarized in Guillemette et al., 2004).

UGT1A8 and UGT2B7 were immunohistochemically determined in epithelium in normal human mammary gland tissue (Thibaudeau et al., 2006). Strikingly, mRNA of UGT1A10 was detected in 19 of 20 normal mammary gland tissue samples (Caucasians) using reverse-transcriptase PCR in Starlard-Davenport et al. (2008) but mRNA was not detected in one of one sample (Caucasian) using reverse-transcriptase PCR in Ohno and Nakajin (2009). Likewise, mRNA of UGT1A3 and UGT1A4 were detected in one of one sample normal breast tissue RNA using reverse-transcriptase PCR in Chouinard et al. (2006), but mRNA was not detected in one of one sample using reverse-transcriptase PCR in Ohno and Nakajin (2009). mRNA of UGT1A9 was not detected in normal breast tissue, based on one sample per study (Ohno and Nakajin, 2009; Lehmann and Wagner, 2008). However, it was immunohistochemically detected in epithelial cells of normal breast tissue (Thibaudeau et al., 2006), but it must be noted that the antibody used in this study was reactive against UGT1A9 and UGT1A8.

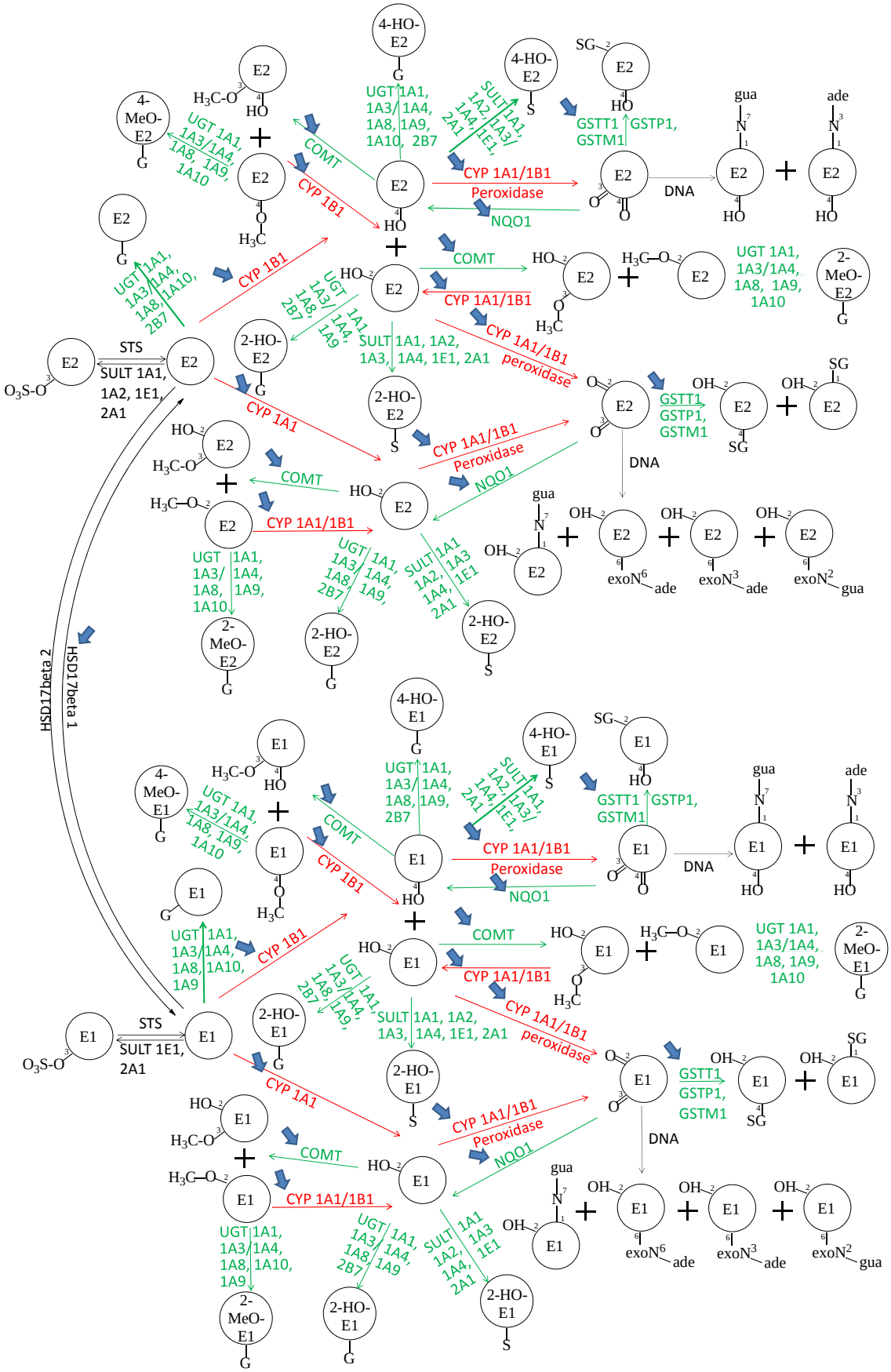


Figure 6: Metabolism of E2 and E1 in normal breast tissue: Activation to genotoxic metabolites (red) and detoxification of genotoxic metabolites (green). Blue arrows indicated polymorphisms known to alter enzyme activity and/or associated with increased breast cancer risk.

Likewise, mRNA of UGT1A1 was not detected in normal breast tissue (Chouinard et al., 2006; Ohno and Nakajin, 2009; Lehmann and Wagner, 2008). However, all mentioned studies, except for Thibaudeau et al. 2006 (n=5), investigating UGT expression in normal breast tissue were only carried out in one sample.

Further detoxification of catechols and quinones

Up to now, only GST-pi (P) has been tested for its ability to metabolize E1- and E2-quinones (Hachey et al., 2003), no data were found for GST-alpha (A), GST-mu (M) or GST-theta (T). All GST isoenzymes (GSTA, GSTM, GSTT and GSTP) were detected in normal breast tissue, where GSTP and GSTT were the most abundant ones (Cairns et al., 1992; Oguztuzun et al., 2011; Hachey et al., 2003; Howie et al., 1990). GSTA and GSTM due to a copy number polymorphism (Rudolph et al., 2012) were only expressed in a few women and in only low amounts (Cairns et al., 1992; Oguztuzun et al., 2011; Hachey et al., 2003; Howie et al., 1990).

Many studies showing that catechol-*O*-methyltransferase (COMT) and NQO1 are metabolizing E1/E2-catechols and -quinones, respectively (summarized in Blair, 2010). COMT catalyzes the methylation of 2-HO-E1/E2 and 4-HO-E1/E2 at position 2 or 3 and 3 or 4, respectively (Dawling et al., 2001; Goodman et al., 2002). No dimethyl ethers were detected (Dawling et al., 2001). Strikingly, methoxyestrogens exert feedback inhibition on CYP1A1 and CYP1B1 mediated oxidative estrogen metabolism (Dawling et al., 2003). NQO1 reduces the E1/E2-quinones to the respective E1/E2-catechols (summarized in Blair, 2010). Both COMT and NQO1 are expressed in epithelial cells of normal mammary gland tissue (Weisz et al., 2000; Siegel and Ross, 2000, respectively).

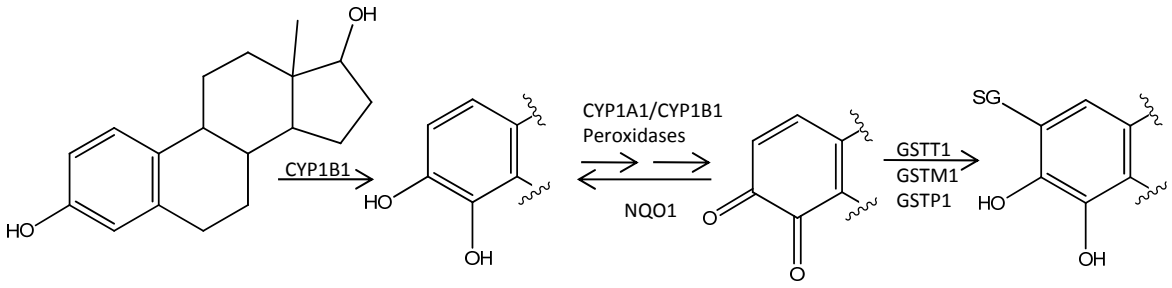


Figure 7: Simplified metabolism of E2 to GSH adducts in normal breast tissue by taking the example of 4-HO-E2.

2.2.3 Transcription factors

There are four main transcription factors known to alter the expression of the enzymes involved in E1/E2 metabolism:

- The nuclear receptor-sub family 1-group I-member 2 (NR1I2, also known as pregnane X receptor, PXR)
- The nuclear receptor-sub family 1-group I-member 3 (NR1I3, also known as constitutive androstane receptor, CAR)
- The nuclear aryl hydrocarbon receptor (AHR)
- The aryl hydrocarbon receptor nuclear translocator (ARNT)

The transcript levels of these transcription factors may also be influenced by age, reproductive and lifestyle factors and thereby also influencing E2 metabolism.

NR1I2

NR1I2 targets members of the CYP3A-family (Qiao et al., 2013) and CYP2C-family (summarized in Smutny et al., 2013) as well as UGT1A1, 1A3, 1A4, 1A6, 1A9 (Qiao et al., 2013), SULT2A1, 1E1 (Kodama and Negishi, 2013), NQO1, GST1A1 and GST1A2 (Kohle and Bock, 2009).

The structure of PXR ligands is commonly hydrophobic with several polar groups, which include xenobiotic and steroid ligands (summarized in Smutny et al., 2013). Additionally, knockout experiments showed the involvement of NR1I2 in regulation of endogenous steroid metabolism (summarized in Zhou et al., 2009; Kodama and Negishi, 2013), which comprised activation of PXR by several endogenous steroids, including E1 and E2 (Banerjee et al., 2013). The activity (including subcellular localization, dimerization and DNA binding) of the PXR can be influenced by post-translational modification, e.g. phosphorylation (Smutny et al., 2013).

The mRNA of NR1I2 has not been detected in mammary gland, however the study included only one sample of mammary gland tissue (Miki et al., 2005).

NR1I3

NR1I3 also targets CYP3A4, 2B6, 2C9, UGT1A1 and 1A9 (Kohle and Bock, 2009) and SULT2A1 and 1E1 (Kodama and Negishi, 2013), as well as GSTs (Yang and Wang, 2014).

CAR activators are clinically used drugs, environmental chemicals and endogenous steroids, like E2, androstenol and androstanol (summarized in Molnar et al., 2013). However, most of these activators do not bind directly to CAR, they activate CAR by stimulating its nuclear translocation in a ligand-independent manner (summarized in Yang and Wang, 2014).

Until now, no data has been published concerning the expression of NR1I3 in normal or malignant mammary gland tissue.

AHR and ARNT

AHR as heterodimer with ARNT binds to xenobiotic responsive elements (Feng et al., 2013) and induces CYPs, including CYP1A1, CYP1A2, CYP1B1 and phase II enzymes

(e.g. UGT1A6), NQO1 and several GSTs, including GSTA1 and GSTA2 (Kohle and Bock, 2009).

AHR ligands are polycyclic aromatic hydrocarbons, polychlorinated biphenyls and aromatic amines (summarized in Feng et al., 2013). AHR activity is negatively regulated by the presence of the AHR repressor protein, whose expression in turn is regulated by AHR (summarized in Feng et al., 2013).

Activated AHR can interact with ESR1/2 activity by various mechanisms (summarized in Feng et al., 2013): ESR1 may co-activate AHR and thus induce CYP1A1 and CYP1B1 expression. Furthermore, AHR ligands induce AHR-mediated ubiquitination and proteasomal degradation of ESR1 and thus suppress estrogen signaling. In addition, AHR may impact the function of ESRs via components of the AHR complex (such as HSP90 and ARNT). ARNT is also found to be a potent co-activator of the ESR signaling (summarized in Feng et al., 2013). Both, ARNT and AHR are expressed in breast cancer cell line MCF-7 (Hanioka et al., 2012). Concerning the expression of AHR in normal mammary gland tissue, only data in normal epithelial cell lines derived from reduction mammoplasty have been published (Yang et al., 2008).

2.3 Cellular stress

Catechol estrogens cause oxidative DNA damage in cultured cells (Mobley et al., 1999). Furthermore, as mechanism of body mass index (BMI)-associated breast cancer risk an increase in proinflammatory mediators in obese women is discussed (summarized in Crujeiras et al., 2013). Inflammation in turn leads to reactive oxygen species which are known to damage DNA (Chapter 2.4).

Most of the genes encoding enzymes involved in cellular stress response contain an antioxidant responsive element (ARE) in their promoter region (summarized in Niture et al., 2010), e.g. γ-glutamyl-cysteine ligase (GCLC, Lee et al., 2003; Hayes and Pulford, 1995; Jeyapaul and Jaiswal, 2000; Lu, 2008). GCLC is the key enzyme of the synthesis of GSH, one of the most important antioxidants in the cell (Richman and Meister, 1975).

Moreover, transcription of some enzymes of phase II metabolism (e.g. NQO1 and GSTs), which are also involved in detoxification of reactive estrogen metabolites, is induced by oxidative stress by binding of the nuclear factor (erythroid-derived 2)-like 2 (NFE2L2) to ARE in their promoter region (Xu et al., 2005, Figure 8). Therefore, higher transcript levels of these enzymes indicate cellular stress and thus are used as marker transcripts for cellular stress (Rushmore et al., 1991).

A marker for DNA damage and other stress signals associated with growth arrest and apoptosis is the growth arrest and DNA-damage-inducible protein 45 alpha (GADD45A,

summarized in Salvador et al., 2013, Figure 8). 14-3-3σ (SFN) is a negative regulator of the cell cycle and it is induced by TP53 to initiate cell cycle checkpoint control after DNA damage, but the mechanism is unclear (summarized in Lee and Lozano, 2006). SFN was immunohistologically detected in MECs of mammary gland tissue (Nakajima et al., 2003).

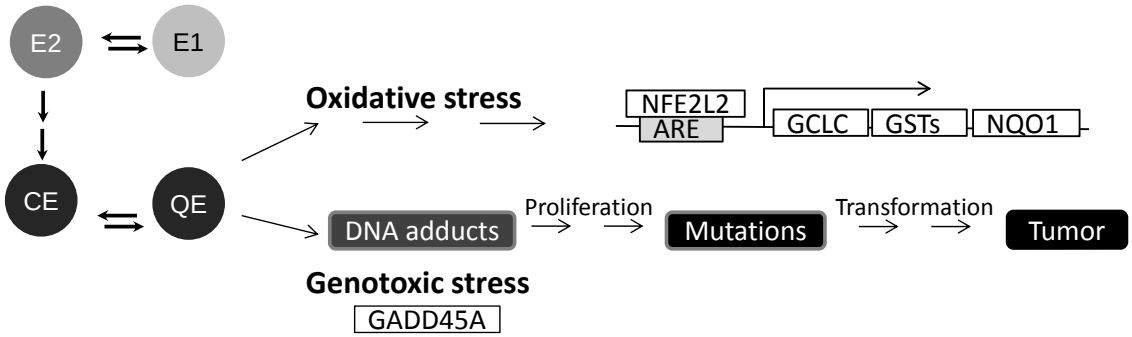


Figure 8: Simplified scheme of cellular stress caused by estrogens in human mammary gland. CE, catechol estrogens. QE, estrogen quinones.

Besides oxidative stress caused by estrogen catechols, both the E1/E2-2,3-quinones and E1/E2-3,4-quinones react with DNA to form DNA adducts (summarized in Cavalieri and Rogan, 2011, Figure 6). The depurinating adducts (4-OH-E1/E2-1-N3 Adenine and 4-OH-E1/E2-1-N32 N7 Guanine, Figure 6) lead to apurinic sites (summarized in Cavalieri and Rogan, 2011), which in turn, with an error-prone DNA repair, can lead to mutations (Cavalieri et al., 2006). E2-enhanced cell proliferation can lead to proliferation of these initiated cells (summarized in Cavalieri and Rogan, 2011).

Hence, in order to determine cellular stress within tissue, besides the quantification of transcripts of cellular stress markers, DNA adducts (Chapter 2.3.1) and mutations (Chapter 2.3.2) can be used.

2.3.1 Metabolic networks

The amount of E2 and its activation to genotoxic metabolites in mammary gland tissue is dependent on interaction of many enzymes, including activation and deactivation of reactive metabolites, therefore the genotoxic stress induced by E2 can only be predicted by appropriate statistical or bioinformatical methods. Statistical methods like principal component analysis or linear regression are able to identify central elements in the metabolism, but they are not able to predict the impact of different transcript level and polymorphism scenarios on whole E1/E2 mediated genotoxicity by calculating DNA adduct formation.

Network analyses are widely used mainly in biological and medical questions, investigating the dynamic behavior of processes by means of simulation to predict response of the biological system (Rezola et al., 2014).

Network analysis is based on computational tools which allows to consider network effects besides measured target effects by complementing the input data by the network modeling (Cecil et al., 2011). As input various data can be used, from mRNA level to enzyme activities. Metabolic processes are typically organized into metabolic pathways, which are commonly defined as a set of consecutive enzyme-catalyzed reactions that convert a set of source/target metabolites. Individual pathways can be found in biochemistry textbooks and several databases, e.g. Kyoto Encyclopedia of Genes and Genomes (KEGG).

One mathematical approach is the constraint-based modeling based on a stoichiometric matrix. The stoichiometric matrix contains all stoichiometric coefficients of the whole network, where every column reflects one reaction and every row one metabolite. The constrain-based modeling has two major constraints: The first is the steady-state condition, i.e. no accumulation or depletion of the metabolites inside the network (summarized in Rezola et al., 2014). The more reactions a metabolite is involved in, the more active is the synthesis reaction. The second is the thermodynamic feasibility, which restricts some fluxes to being non-negative, because of their associated Gibbs free energy (summarized in Rezola et al., 2014). Together with the input data (e.g transcript level for each enzyme) the model calculates the resulting fluxes within the network based on a third constraint, the non-decomposability condition, which ensures that the calculated solution comprises a minimal number of active reactions at steady-state. This implies that these solutions cannot be decomposed into any smaller flux distributions without violating the steady-state constraint (summarized in Rezola et al., 2014). The resulting fluxes describe the activity of the respective reaction in the whole network even for reactions where input data had been available. Thus for example, a prediction on DNA-adduct formation can be used as a marker for the impact of the transcript level of enzymes involved in E1/E2 metabolism and polymorphisms on the whole E1/E2 metabolism.

2.3.2 Determination of spontaneous mutation frequency and mutation spectra

Another approach to estimate E2 mediated genotoxicity is to determine the mutations directly in the mammary gland tissue.

Data on SMF in intron sequences were very rare, the most recent one was published in FEN1 gene in normal human lung tissue ($3x10^{-7}$, Zheng et al., 2007), older publications in normal skin fibroblasts determined a SMF in intron sequence of the TP53 gene of $1x10^{-8}$ (Bielas and Loeb, 2005).

It is expected, that the spontaneous mutations frequency is very low, down to $1x10^{-8}$ per base pair in intron sequences (Bielas and Loeb, 2005; Zheng et al., 2007), therefore a method is needed which is able to detect this mutation frequency, without previous clonal expansion, since mammary gland tissue is not divisible *ex vivo*. The Random Mutation Capture Assay

(RMCA) is an extremely sensitive genotypic selection method for quantifying the frequency of random mutations and was already applied to cultured normal skin fibroblasts (Bielas and Loeb, 2005). Thus, this method could be appropriate for the detection of the SMF in normal mammary gland tissue. Mutations within the mutation target are determined by digesting double-stranded DNA with the restriction enzyme *Taq*I. It cleaves the wild-type sequence (*TCGA*) but will not digest the DNA if a mutation is present within this recognition site. Because of the presence of the *Taq*I recognition sequence in the intron and exon of most genes the RMCA is widely applicable.

The RMCA includes nine steps (Figure 9):

1. Selection of sequences of target genes containing one *Taq*I recognition site.

2. Digestion of genomic DNA with restriction enzymes that do not target the chosen sequence containing *Taq*I recognition site.

3. Hybridization of DNA with a biotinylated DNA probe, containing uracil, specific for the selected sequence.

4. Enrichment of the hybridized target by magnetic separation after complexing with streptavidin coupled to superparamagnetic polymer spheres.

5. Digestion of DNA with *Taq*I. If there are any mutations in the *Taq*I recognition site (TCGA) the sequence will not be cleaved.

6. DNA probe is removed by digestion with uracil-DNA glycosylase, abolishing its ability to serve as template in PCR.

7. Quantitative PCR to determine the total number of target sequences.

8. PCR to determine mutant frequency.

9. Verification of mutations by resistance to *Taq*I digestion. Moreover, DNA sequencing provides the mutation spectra.

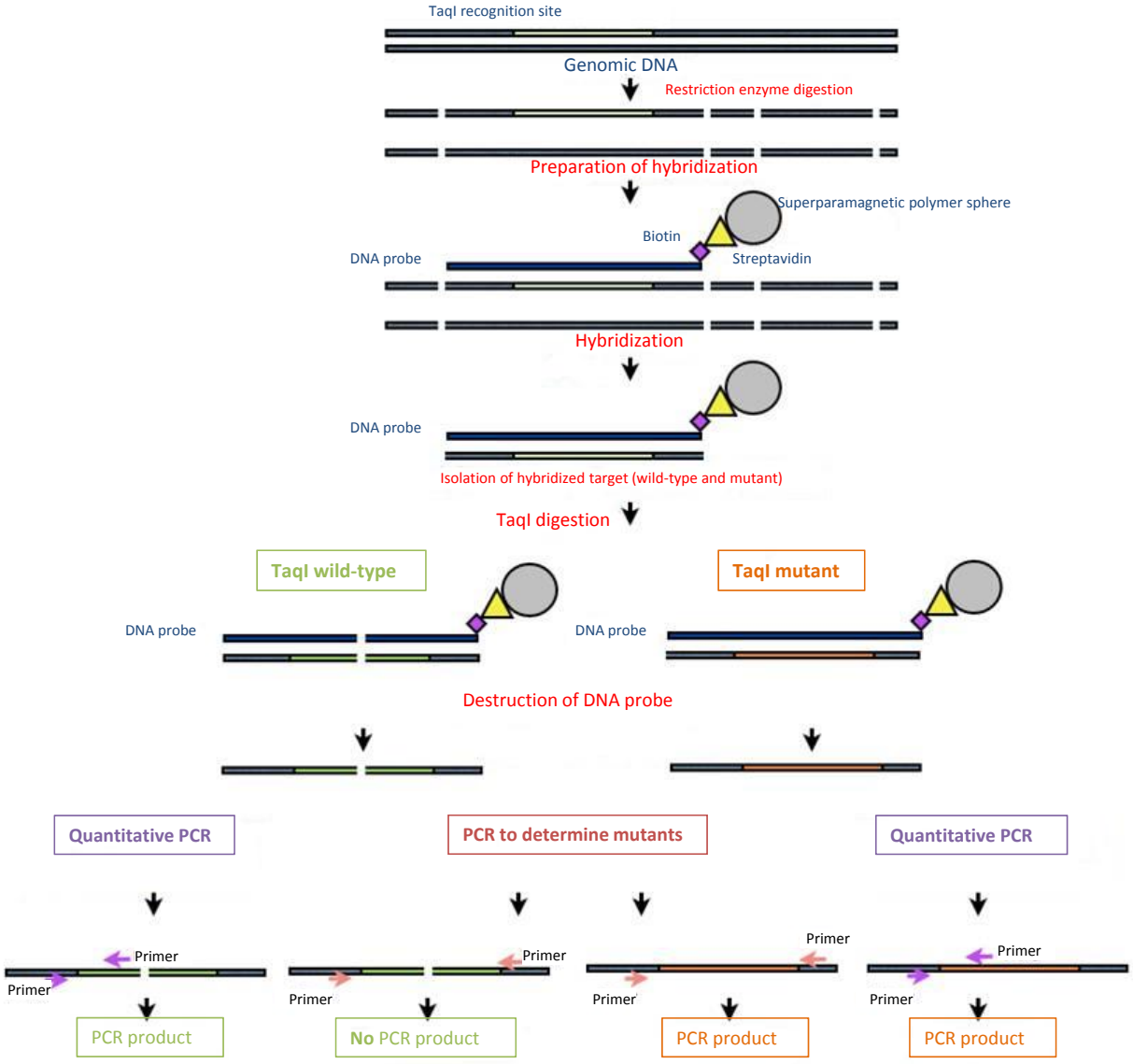


Figure 9: Scheme of RMCA (according to Bielas and Loeb, 2005).

2.4 Factors influencing breast cancer risk by interaction with E2 biosynthesis and metabolism

Some factors were identified in epidemiological studies influencing breast cancer risk in Caucasian women, such as BMI, parity, smoking and alcohol consumption, which may act by interfering with E1/E2 biosynthesis and metabolism (summarized in Coronado et al., 2011; Key et al., 2011). Likewise, some polymorphisms of enzymes involved in E2 metabolism are epidemiologically correlated with a higher breast cancer risk in Caucasian women (Zhang et al., 2011, Chapter 2.4.1), which may define subpopulations of women which have a higher endogenous exposure to reactive E2 metabolites and may thus cause a higher susceptibility

for developing breast cancer (Zhang et al., 2011), because of increased genetic damage during lifetime (Chapter 2.4.1).

BMI

Obesity is epidemiologically associated with an increased breast cancer risk in postmenopausal women (summarized in Rose and Vona-Davis, 2010): E2 can be synthesized in adipose tissue by CYP19A1. Thus, biologically available E2 is elevated in the circulation of obese women which is discussed as promotional effect on postmenopausal breast carcinogenesis (summarized in Rose and Vona-Davis, 2010). Furthermore, it was recently found that the CYP19A1 activity in adipose tissue is paralleled with an increase in proinflammatory mediators, such as proinflammatory cytokines, in obese women (summarized in Crujeiras et al., 2013). Inflammation in turn leads to reactive oxygen species which are known to damage DNA (summarized in Crujeiras et al., 2013).

Smoking and alcohol

Smoking and alcohol consumption are also epidemiologically associated with an increased breast cancer risk, but the mechanisms are unclear (summarized in Coronado et al., 2011; Gaudet et al., 2013). Elevated circulating sex hormone plasma levels in postmenopausal women were found in smokers compared to non smokers, which is discussed as a mechanism for the increased breast cancer risk (Key et al., 2011). A lower serum E1 level has been related to smoking by nicotine increasing the number of regressing follicles in the ovary and inhibiting CYP19A1, which leads to a decreased conversion of androgens to estrogens (summarized in Tanko and Christiansen, 2004).

Increased plasma level of E1 sulfate and DHEA were detected in alcohol consumers compared to non-alcohol consumers in a controlled nutrition study of 51 postmenopausal women (summarized in Coronado et al., 2011). Furthermore, genotoxic metabolites of ethanol or reactive oxygen species, induced by ethanol *in vitro* are discussed as reasons for the increase of breast cancer risk (summarized in Coronado et al., 2011).

Parity

Full time pregnancy is known to reduce the risk of breast cancer (Reeves et al., 2009), but the molecular mechanisms are still unclear. It was observed in animal models, that short term treatment with pregnancy levels of E2 conferred protection against N-methyl-N-nitrosourea-induced breast cancer in mice (Rajkumar et al., 2001), mimicking the protective effect of parity, by targeting major cell signaling pathways involved in cell survival, cell migration and cell death (Arumugam et al., 2014).

2.4.1 Polymorphisms in genes involved in E2 metabolism

Individual differences in the metabolism of E2 in breast tissue might have a significant impact on the development of breast cancer. An enhanced activity in E2 catechol and quinone-

forming and a reduction in detoxifying enzymes could affect the estrogen toxicity in breast tissue and thus favor the formation of breast cancer (Singh et al., 2005). Many different polymorphisms in both E2 toxifying enzymes, such as CYP1A1 and 1B1 and detoxifying enzymes such as NQO1, COMT and GSTs (Jiao et al., 2010; Mitrunen and Hirvonen, 2003; Miyoshi and Noguchi, 2003; Guillemette et al., 2004; Huang et al., 1999) have been investigated in epidemiological studies. Not all studies yielded uniform results. A recent review summarized all meta-analysis for polymorphisms of enzymes involved in E1/E2 metabolism (Zhang et al., 2011). They found GSTT1 copy number polymorphism (CNP), GSTM1 CNP and single nucleotide polymorphisms (SNPs) NQO1 Pro187Ser (rs1800566), HSD17B1 (rs676387) as significant associated with breast cancer in Caucasians (Figure 6). When using the dominant or recessive models additionally the polymorphism CYP1B1 Leu432Val (rs1056836) was significantly associated with breast cancer risk (Figure 6).

The COMT Val108/158Met (rs4680) polymorphism is not correlated with an increased risk of breast cancer (Zhang et al., 2011), although the low activity allele (Met/Met) has significantly reduced activity against E2 catechols (Dawling et al., 2001). In vitro studies show that simultaneous lack of NQO1 and COMT activity results in increased DNA adduct formation in comparison to NQO1-deficient cells with normal COMT activity (Singh et al., 2009). Thus, women who have both genotypes COMT Met/Met and NQO1 Ser/Ser, could have an increased estrogen mediated genotoxicity. However, an epidemiological study investigating the interaction of COMT Val108/158Met and NQO1 Pro187Ser has been lacking until now.

2.4.1.1 NQO Pro187Ser SNP

NQO1 is involved in detoxifying highly reactive quinone metabolites of E2 and E1 (Figure 6). It catalyzes the reduction of the quinones to the corresponding catechols (Chapter 2.2.2). Thus, NQO1 also plays a central role in protecting cells from oxidative stress. The exchange of a cytosine with a thymine at position 609 in the transcript of the gene, results in replacement of Pro to Ser in amino acid sequence at position 187 (Table 3). The NQO1 Ser/Ser protein was shown to have reduced activity in cancer cell lines (Nebert et al., 2002; Siegel et al., 1999, 2001). The homozygous variant (Ser/Ser) has only 2% of the activity of the homozygous variant (Pro/Pro, Traver et al., 1997). This reduction in activity is based on a shorter half-life time of the Ser/Ser protein (1.2 h) compared to the Pro/Pro protein (18 h, Siegel et al., 2001).

Table 3: Summary of polymorphisms investigated in this work.

Enzyme	Change in DNA sequence (position in mRNA)	Amino acid change (position)	rs number	Genetic variant	Frequency N (%) among German Caucasians	References
COMT	G to A (452, cytosolic form; 696, membrane bound form)	Val to Met (108 cytosolic form, 158 membrane bound form)	rs4680	Val/Val Val/Met Met/Met	22.0-28.0 49.0-54.9 24.0-28.0	Peterson et al. 2010; Mao et al. 2010a; Rudolph et al. 2011; Schmahl et al. 2012; Reuter et al. 2006; Majic et al. 2011
NQO1	C to T (609)	Pro to Ser (187)	rs1800566	Pro/Pro Pro/Ser Ser/Ser	67.6-73.4 25.0-32.4 1.6	Nebert et al. 2002; Yuan et al. 2011; Singh et al. 2009; Timofeeva et al. 2010; Zhang et al. 2003
GSTT	Deletion	–	–	+/+ +/- -/-	32.6-33.5 49.4-50.0 17.1-19.3	Dong et al. 2008; Sergentanis and Economopoulos 2010; Timofeeva et al. 2010; Bruhn et al. 1998; Rudolph et al. 2012

For the NQO1 polymorphisms Pro187Ser an increased breast cancer risk (27%, Zhang et al. 2011) has been identified for Caucasian women with Ser allele (Ser/Ser vs. Pro/Pro and Ser/Ser + Ser/Pro vs. Pro/Pro, Yuan et al. 2011). Thus, the risk is also increased for the heterozygous genotyp Pro/Ser. The frequency in Caucasian women with German ancestry ranges from 67.6 to 73.4% Pro/Pro, 26.6 to 32.4% Pro/Ser and 1.6% Ser/Ser (Table 3).

2.4.1.2 GSTT1 CNP

The GSTs are, besides NQO1, important enzymes for the detoxification of reactive E1(E2)-quinones (Figure 6). The GSTT1 CNP is based on the total or partial deletion of both (nullizygous or -/-) or one copy (hemizygous or +/-) of the gene (Pemble et al., 1994, Table 3). The nullizygous genotype of GSTT1 has no enzyme activity (Pemble et al., 1994) and is associated with an increase in breast cancer risk of 11% (Zhang et al., 2011). Among German Caucasians GSTT1 -/- occurs in 17.1-19.3%, +/- in 49.4-50.0% and +/+ in 32.6-33.5% of the population (Table 3).

2.4.1.3 COMT Val108/158Met SNP

COMT catalyzes the methylation of catechol estrogens, which then can be further detoxified by glucuronidation or sulfonation (Figure 6). The COMT Val108/158Met polymorphism is based on an exchange of guanine with adenine at position 452 of the cytosolic form and at position 696 of the membrane bound form in the transcript. This results in replacement of Val to Met at position 108 of the cytosolic form and 158 of the membrane bound form (Table 3). The COMT Met/Met protein has only approx. 50% of the activity of the Val/Val-protein, the mRNA level is not influenced (Dawling et al., 2001; Chen et al., 2004). The difference in activity is based on a better thermostability of the Val/Val protein (Syvanen et al., 1997). The outcomes of studies assessing the association of COMT Val108/158Met polymorphism and breast cancer risk are inconclusive (Ding et al., 2010; He et al., 2012; Huang et al., 1999). Recent meta-analysis, however, comes to the conclusion that there is no correlation between the low-activity COMT Met/Met variant and breast cancer risk (Zhang et al., 2011). The frequency among the German population is 24.0-28.0% Met/Met, 49.0-54.9% Val/Met and 22.0-28.0% Val/Val (Table 3).

3 Objectives

Breast cancer is the most common cancer among women in Germany. Only a small share of breast cancer patients have heritable mutations and therefore an increased risk of developing breast cancer, while the rest of them may acquire spontaneously genetic mutations during their lifetime. The female sex hormone 17β-estradiol, produced naturally in the body, seems to play an important role in the development of breast cancer: (i) 17β-estradiol can be activated to reactive metabolites by enzymes of the cytochrome P450 family present in mammary gland tissue. These reactive metabolites are known to damage the DNA. (ii) The stimulation of the estrogen receptor α by 17β-estradiol enhances cell proliferation which can indirectly increase the mutation frequency.

There are factors affecting breast cancer risk by presumably modulating 17β-estradiol metabolism: Polymorphisms in 17β-estradiol metabolizing enzymes change the activity or amount of specific enzymes which in turn influences the whole metabolism. Likewise, lifestyle and environmental factors, which are epidemiologically correlated with an higher breast cancer risk, may influence the expression of enzymes involved in 17β-estradiol metabolism.

To investigate the causes of breast cancer it is important to understand processes within the tissue prior to transformation (normal tissue). However, the influence of these factors, affecting breast cancer risk, on 17β-estradiol metabolism in its entirety in normal mammary gland tissue has not been investigated until now.

Therefore, the aim of this work is to investigate the influence of polymorphisms in genes encoding 17β-estradiol metabolizing enzymes and lifestyle factors on tissue levels of 17β-estradiol and estrone as well as on genotoxic stress and proliferation/differentiation within mammary gland tissue.

First, the factors associated with breast cancer risk and possible influence on 17β-estradiol metabolism as well as polymorphisms in genes encoding enzymes of 17β-estradiol metabolism, correlated with higher breast cancer risk, will be identified.

Mammary gland tissue of healthy women will be collected from women who undergo mamma reduction surgery for cosmetic reasons. Out of reduction surgery a large quantity of very inhomogeneous breast tissue will collected, thus the preparation needs to be standardized in order to achieve better comparability among the individual samples. Characterization of tissue concerning the factors associated with breast cancer risk, will be performed by questionnaire (lifestyle factors), genotyping and histology. Since the methods have not been established in the work group, they will be established and their performance evaluated.

In order to investigate the influence of factors associated with breast cancer risk on 17β-estradiol metabolism, important (iso)enzymes in mammary gland E2 metabolism considering presence in mammary gland tissue and involvement in 17β-estradiol metabolism will be

identified. Furthermore, transcripts of genes suitable for the characterization of tissue concerning proliferation, differentiation, apoptosis and paracrine signal transduction in normal mammary gland tissue will be identified.

To quantify and characterize transcription levels of all important transcripts a quantitative method based on low density TaqMan™-probe based array-technique will be established.

Since some transcript are expected to be expressed very low, suitability of preamplification reaction prior to real time PCR will be tested.

In order to describe the differences in 17β-estradiol metabolism of normal human mammary gland tissue and the human breast cancer cell line MCF-7, often used for studying influencing factors on breast cancer risk, the levels of transcripts encoding enzymes of 17β-estradiol metabolism in normal tissue will be compared to the respective transcript levels in MCF-7 cells.

Since epithelial cells are the origin of most of the breast tumors, the putative gain of information by analyzing epithelial cells isolated by laser microdissection compared to mammary gland homogenate will be investigated. The extent of genotoxic stress within the mammary gland tissues will be assessed by transcript levels of marker genes and DNA adduct fluxes calculated by bioinformatical network analysis. Furthermore, the suitability of the Random Mutation Capture assay to determine frequencies of spontaneous mutations will be investigated.

Finally, to investigate the influence of polymorphisms and lifestyle factors on 17β-estradiol mediated genotoxicity and to answer the question if the 17β-estradiol metabolism is decisive for the 17β-estradiol mammary gland tissue levels, appropriate statistical methods will be selected and applied.

4 Problemstellung

Brustkrebs ist die häufigste Krebserkrankung bei Frauen in Deutschland. Nur ein kleiner Anteil der Brustkrebspatientinnen hat vererbte Mutationen und damit ein erhöhtes Risiko an Brustkrebs zu erkranken. Die übrigen von ihnen erwerben Mutationen spontan im Laufe ihres Lebens. Das weibliche Sexualhormon 17β-Estradiol, das natürlich im Körper synthetisiert wird, scheint ursächlich an der Entwicklung von Brustkrebs beteiligt zu sein: (i) 17β-Estradiol kann durch, in der Brustdrüse vorhandene, Enzyme der Cytochrom-P450-Familie zu reaktiven Metaboliten aktiviert werden, welche dann die DNA schädigen können. (ii) Stimulierung des Estrogenrezeptors α durch 17β-Estradiol erhöht die Zellproliferation, und somit indirekt die Mutationsfrequenz und trägt daher zur Tumorpromotion und -progression bei.

Die meisten Faktoren, die mit einem erhöhten Brustkrebsrisiko assoziiert sind, wirken vermutlich durch Modulation des 17β-Estradiol-Metabolismus: Polymorphismen in 17β-Estradiol-metabolisierenden Enzymen können die Aktivität oder Menge dieser Enzyme verändern. Ebenso können Lebensweise und Umweltfaktoren die Expression der 17β-Estradiol-metabolisierenden beeinflussen.

Um die Ursachen der Initiation von Brustkrebs im histologisch normalen Gewebe zu untersuchen, ist es wichtig die Auswirkungen dieser Faktoren auf das Gewebe vor einer bösartigen Transformation der Zellen zu kennen. Trotz seiner großen Bedeutung ist der Einfluss dieser Faktoren auf den gesamten 17β-Estradiol-Metabolismus in normalen Brustdrüsengewebe bisher nicht untersucht worden.

Deshalb war das Ziel dieser Arbeit den Einfluss von Lebensweise und Polymorphismen auf genetoxischen Stress und Prolifertaion/Differenzierung im Brustdrüsengewebe zu untersuchen.

Zunächst werden die Faktoren, die mit einem erhöhten Brustkrebsrisiko assoziiert sind und möglicherweise einen Einfluss auf den 17β-Estradiol-Metabolismus haben, sowie Polymorphismen in Genen, die für 17β-Estradiol-metabolisierende Enzyme kodieren und mit einem erhöhten Brustkrebsrisiko korreliert werden, identifiziert.

Brustdrüsengewebe von gesunden Frauen, die sich aus kosmetische Gründen einer Mammareduktion unterziehen, werden gesammelt. Da das entnommene Brustgewebe sehr inhomogen ist, soll die Präparation standardisiert werden, um die Vergleichbarkeit zwischen den einzelnen Proben zu gewährleisten. Die Charakterisierung des Gewebes hinsichtlich der Faktoren, die mit einem erhöhten Brustkrebsrisiko assoziiert sind, soll unter Verwendung von Fragebögen (Lebensweise), Genotypisierung von Polymorphismen und der Histologie vorgenommen werden. Da diese Methoden am Lehstuhl bisher nicht etabliert sind, sollen sie etabliert werden und hinsichtlich ihrer Leistungsstärke charakterisiert werden.

Um nun den Einfluss der Faktoren, die mit einem erhöhten Brustkrebsrisiko assoziiert sind, auf den 17β-Estradiol-Metabolismus zu untersuchen, sollen wichtige (Iso)-Enzyme des Metabolismus in der Brustdrüse identifiziert werden. Darüber hinaus sollen Transkripte der Gene identifiziert werden, die geeignet sind, um sowohl die proliferation-steigernden Eigenschaften von 17β-Estradiol als auch die Differenzierung, Apoptose und die parakrine Signalübertragung in normalem Brustdrüsengewebe zu charakterisieren.

Um die Transkriptlevel aller wichtigen Transkripte zu quantifizieren, soll eine Echtzeit-TaqMan™-Sonden basierte Polymerasekettenreaktionsmethode etabliert und ihre Leistungsfähigkeit charakterisiert werden. Da einige Transkriptlevel sehr niedrig erwartet werden, soll die Eignung einer Prä-amplifizierung vor der Echtzeit-Polymerasekettenreaktion überprüft werden.

Um die Unterschiede im 17β-Estradiol-Metabolismus zwischen dem normalen menschlichen Brustdrüsengewebe und einer humanen Brustkrebszelllinie (MCF-7), die oft für Studien zur Untersuchung von Einflussfaktoren auf das Brustkrebsrisiko benutzt werden, zu zeigen, sollen die Transkriptlevel, die für Enzyme des 17β-Estradiol-Metabolismus kodieren, von MCF-7 Zellen und normalem Brustdrüsengewebe verglichen werden.

Da die Epithelzellen der Ursprung der meisten Brusttumoren sind, soll der vermeintliche Informationsgewinn durch die Analyse von Epithelzellen, die durch Laser-Mikrodissektion isoliert werden, im Vergleich zu Brustdrüsenhomogenat untersucht werden.

Die TaqMan™-Sonden basierte Methode soll auch zur Bestimmung der Transkriptlevel in Laser-Mikrodissektierten Epithelzellen der Brustdrüsen angewendet werden.

Das Ausmaß des genotoxischen Stresses innerhalb des Brustdrüsengewebes soll mittels Transkriptleveln von Marker-Genen und DNA-Addukt Flüssen, welche durch bioinformatische Netzwerkanalyse berechnet werden, beurteilt werden. Ferner soll die Eignung des *Random Mutation Capture*-Assays zur Bestimmung der spontanen Mutationsfrequenz im Brustdrüsengewebe untersucht werden.

Um den Einfluss von Polymorphismen und der Lebensweise auf 17β-Estradiol-vermittelte Genotoxizität zu untersuchen und um die Frage zu beantworten, ob der gewebseigene 17β-Estradiol-Metabolismus entscheidend für die 17β-Estradiol-Gewebespiegel im Brustdrüsengewebe ist, sollen geeignete statistische Methoden ausgewählt und angewendet werden.

5 Materials and Methods

5.1 Materials

5.1.1 Equipment and laboratory consumables

Equipment	Name (Provider)
Autoclave	Systec DX-90 2D (Systec GmbH, Wettenberg, Germany)
Benchtop centrifuge	Combi Spin FVL. 2400 Vortex (PEQLAB Biotechnologie GmbH, Erlangen, Germany)
	Perfect Spin 24 (PEQLAB Biotechnologie GmbH)
Cell counting device	Electronic CASYR Model DT (Schärfe System, Reutlingen, Germany)
Centrifuges	Benchtop centrifuge: Combi-Spin, Part No.: FVL-2400 N with Vortex (PEQLAB Biotechnologie GmbH)
	Small centrifuge: EBA 12 (Andreas Hettich GmbH & Co.KG, Tuttlingen, Germany)
	Centrifuge for 96-well plates: Heraeus Multifuge 3SR+ (Thermo Fisher Scientific Inc., Waltham, Germany)
Cryostat	CM1950 (Leica Microsystems, Wetzlar, Germany)
DNA/RNA quantification plate for TECAN plate reader	NanoQuant Plate™, Part No.: 3003594 (Tecan Deustchland GmbH, Crailsheim, Germany)
Drying oven	Memmert, Schwabach, Germany
Electrophoresis chambers	VG-SYS UNIT COMPLETE (Part No.: 30019564, Carl Roth GmbH & Co. KG, Karlsruhe, Germany)
	PerfectBlue Gelsystem Mini M (Part No.: 40-0911, PEQLAB Biotechnologie GmbH)
	PerfectBlue Gelsystem Mini ExM (Part No.: 40-1410, PEQLAB Biotechnologie GmbH)
	PerfectBlue Gelsystem ExW (Part No.: 40-2314, PEQLAB Biotechnologie GmbH)
	PerfectBlue Gelsystem Maxi S (Part No.: 41-1325, PEQLAB Biotechnologie GmbH)
Geldocumentation system	Molecular Imager®, ChemiDocTM XRS+ (BIO-RAD Laboratories GmbH, München, Germany)
Incubator	CO_2 incubator CB 150 (BINDER, Tuttlingen, Germany)
	HERAcell 150i, 240i (Thermo Scientific, Schwerte, Germany)
Laminar flow	Lamina Nu. 480.600 (IBS Integra Biosciences, Fernwald, Germany)
Laser microdissection system	Leica LMD 6500 incl. Laser (CryLaser, type 2CONBOOSTA: wavelength 355 nm, pulse frequency 80 Hz, pulse length <4 ns, average pulse energy 70 μJ, Leica Microsystems)

Equipment	Name (Provider)
Microscopes	Optical microscope (Hund, Wetzlar, Germany) Optical microscope, 8800 Watt, device number: 3447 (SHARP Electronics GmbH, Hamburg, Germany) Optical microscope, Part No.: DM6000B incl. control box (Part No.: CTR6000) and camera (Part No.: DFC310FX, Leica Microsystems)
Mortar and pestle for frozen tissue	self-construction, Institute of Pharmacy and Food Chemistry, University of Würzburg
MultipetteR plus	Eppendorf AG, Hamburg, Germany)
Objectives	10x/0.25, Ph 1 ADL (Nikon GmbH, Düsseldorf, Germany) UVI 6.3x/0.13 Microdissection, (Part No.: 518145, Leica Microsystems) L 40x/0.60 CORR XT, HCX PL FLUOTAR (Part No.: 506208, Leica Microsystems)
PCR heating block	Primus 96 Thermal Cycler (MWG Biotech, Ebersberg, Germany)
pH meter	inoLabR pH 720 (WTW, Weilheim, Germany)
Photometer	UV mini-1240 (Shimadzu GmbH, Duisburg, Germany) Eppendorf ResearchR 100 μl, 1000 μl, 500 μl, 5000 μl (Eppendorf AG)
Pipettes	0.5 10 μl Discovery Comfort DV10 (Part No.: 4042, ABIMED GmbH, Langenfeld) 10 100 μl Discovery Comfort DV100 (Part No.: 4044, ABIMED GmbH) 100 1000 μl Discovery Comfort DV1000 (Part No.: 4049, ABIMED GmbH) Multipette® plus (Eppendorf AG)
Plate reader	Infinite® 200 PRO series (Tecan, Maennedorf, Switzerland)
Plate reader Tecan GENios	Spektrophotometer Nanodrop ND. 1000, Nanoquant Plate (Part No.: 30035094, Tecan)
Plate stirrer with heating	IKAR RCT basic, (Ikar, Staufen, Germany)
Precision scale	SI-2002, max. 2000 g, d=0.01 g (Denver Instrument GmbH, Göttingen, Germany) SI. 2002 (Denver Instrument)
Real time PCR	7900 HT Fast Real Time PCR System, Instrument Serial No.:279002406, Software: SDS 2.4 (Applied Biosystems, Darmstadt, Germany)
Rotation shaker	Multi Bio RS-24 (Part No.: 320901006, PEQLAB Biotechnologie)
Thermocycler	Primus 96 Thermal Cycler (MWG Biotech)
Thermomixer	Thermomixer compact (Part No.: 5355 000.011, Eppendorf AG)
Ultraviolet Sterilizing PCR Workstation	PEQLAB Biotechnologie GmbH

Equipment	Name (Provider)
UV transilluminator	Model PHER01 40 M, 230 V, 50 Hz (Part No.: 000056520, Biotec-Fischer GmbH, Reiskirchen, Germany)
Vacuum pump	PC 2002 VARIO/ PC 2004 VARIO (Vacuubrand GmbH & Co KG, Wertheim, Germany)
Voltage regulator for gel electrophoresis	POWER PAC 1000, 200/240 V (Part No.: 165-5055, Bio-Rad Laboratories GmbH) WNB10 (Memmert)
Water bath	WB. 4 MS (Biosan, Riga, Latvia) NGW Lauda MS (device number: J16030, Dr. R. Woboser KG, Knigshofen, Germany)
Workstation	Ultraviolet sterilizing PCR workstation, (Part No.: 90-UV/PCR, PEQLAB Biotechnologie GmbH)

5.1.2 Laboratory consumables

Laboratory consumables	Name (Provider)
96 well plates for gel electrophoresis	microtest plate 96-Well, flat bottom (Part No.: 82.1581, Sarstedt AG & Co, Nümbrecht, Germany)
96 well plates for real-time PCR	Micro Amp® Fast Optical 96 well reaction plate with barcode, 0.1 ml (Part No.: 4346906, Applied Biosystems)
Blades	Low Profile Microtome Blades for standard application (Part No.: DB80LS, Leica Microsystems)
CASY® cup	Schärfe System
CASYTON	Schärfe System
Cell culture flasks	Cellstar: 550 ml (175 cm^2 cultivation area), 250 ml (75 cm^2), 50 ml (22 cm^2, Greiner bio-one, Frickenhausen, Germany)
Centrifuge tubes	50 ml conical centrifuge tubes with screw cap (Greiner bio-one) 50 ml conical centrifuge tubes with screw cap (Sarstedt) 15 ml conical centrifuge tubes with screw cap (Sarstedt)
Fast PES Bottle Top Filter,	75 mm diameter of membrane, 45 mm neck, 0.2 μm pore size, NNIItem: 5954520 (Nalgen Nunc International, Roskilde, Denmark)
Glass hook	Glass Pasteur pipette is melted in the flame of a bunsen burner until a hook is formed
Glass Pasteur pipette	Facility for Chemicals and Materials, University of Würzburg
Injection needle	100 Sterican® 0.90 x 40 mm (Part No.: 4657519, B. Braun Melsungen AG, Melsungen, Germany)
Lint free cloths	KIMCARE® 76 x 144 = Code 3020 (Kimberly-Clark Europe Limited/Professional Sector, Surrey, UK)
Magnetic Beads	Dynabeads® kilobaseBINDER™ Kit (Part No.: 60101, Lifetechnologies, Darmstadt, Germany)

Laboratory consumables	Name (Provider)
Membrane slides	PEN-Membrane slides (2.0 μm, Leica Microsystems)
Microcon centrifugal filter	Ultracell YM-50 filters (Part No.: YM-50, Merck KGaA, Darmstadt)
Optical adhesive film for RT PCR	Micro Amp™ Optical Adhesive Film (Part No.: 4311971, Applied Biosystems)
PCR soft tubes	0.2 ml, DNase, RNasefree (Part No.: 710920, Biozym Scientific GmbH, Hessisch Oldendorf, Germany)
Photometer cuvette	Eppendorf UVette®; (1 cm path length; 50 μl sample volume, Eppendorf)
Pipette tips	plastic 1 ml and 200 μl (Sarstedt)
Pipette tips for gel electrophoresis	Pipette tips 10 μl (Part No.: 70.1130, Sarstedt), autoclaved before use Combitips plus 0.1 ml (Part No.: 0030 069.200, Eppendorf AG) Combitips plus 5.0 ml (Part No.: 0030 069.250, Eppendorf AG)
Pipette tips for PCR	SafeSeal Tips Professional 10 μl; 100 μl; 1250 μl, steril Low Binding Technologie, Micro Extent DNA, DNase-, RNase-, Pyrogen-free (Part No.: 770020, 770100, 770600, Biozym)
Plastic Pasteur pipettes	Transfer pipette (Part No.: 861171.001, Sarstedt)
Reaction tubes	Multiply®-pro tube 0.2 ml, PP (Part No.: 72.737.002, Sarstedt) Safe Seal Micro Tube 1.5 ml, PP (Part No.: 72.690.001, Sarstedt) Micro Tube 2 ml, PP (Part No.: 72.695, Sarstedt)
Transferpipette	sterile, 3.5 ml (Part No.: 86.117.001, Sarstedt)

5.1.3 Chemicals and enzymes

Chemicals and enzymes	Name (Provider)
2-Propanol	Isopropanol p.A. (Part No.: 9866.2, Carl Roth GmbH)
Agaroses	peqGOLD Universal, 500 g (Part No.: 35-1020, PEQLAB Biotechnologie GmbH) Agarose (Part No.: A9539-500G, Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)
Calf thymus DNA	
Chloroform	p.A., ISO, stabilized with ethanol (Part No.: 2445.2500, Merck)
Cresyl violet acetate	Cresyl Violet Acetat, certified (Part No.: C5042, Sigma-Aldrich)
Disodium hydrogen phosphate	(Na_2HPO_4), Part No.: p030.1, >99%, p.a.ACS, water-free (Carl Roth GmbH)
DNA ladders	50 bp DNA Ladder (Part No.: N3236L, New England Biolabs GmbH, Frankfurt am Main, Germany) 100 bp DNA Ladder (Part No.: N3231L, New England Biolabs GmbH) GeneRuler™ 1kB DNA ladder (Part No.: SM0311, Thermo Scientific)

Chemicals and enzymes	Name (Provider)
dNTPs	dNTPs, single tubes (dATP/dTTP/dGTP/dCTP) each 100 mM (Part No.: DNTP100-1KT, Sigma-Aldrich) dUTP 100 mM (Part No.: D0184-25UMO, Sigma-Aldrich) dNTP-Mix dATP/dTTP/dGTP/dCTP, 5 mM each (Part No.: 11NTPMX050, MP Biomedicals, Eschwege, Germany)
EDTA	Ethylenediaminetetraacetate disodium, dihydrate (Part No.: 8043.2, Carl Roth GmbH)
Eosin	Eosin Y, dye content ˜99% (Part No.: E4009-5G, Sigma-Aldrich)
Ethanol	Ethanol absolute Molecular biology grade (Part No.: A3678, Applichem GmbH, Darmstadt, Germany)
Glycerol	≥98%, Ph.Eur., water free (Part No.: 7530.1, Carl Roth GmbH)
Loading Dye	Gel Loading Dye, Blue (6X), (Part No.: B7021S, New England Biolabs GmbH) Gel Loading Solution Type I, 6x (Part No.: G7654, Sigma-Aldrich)
Magnesium chloride	$MgCl_2$, 25 mM (Part No.M8787, Sigma-Aldrich)
Sodium bicarbonate	$NaHCO_3$ (Part No.: S57611KG, SigmaAldrich)
Nuclease-free water (ncf-water)	Water Molecular Biology Grade (Part No.: A7398, AppliChem) Water Molecular Biology Grade (Part No.: r058, Thermo Scientific)
PCR buffer	10x PCR Buffer without $MgCl_2$: 100 mM Tris-HCl (pH 8.3), 500 mM KCl (Part No.: P2317, Sigma-Aldrich) Ampliqon 10x TEMPase Buffer II (Part No.: 220302, Biomol) Phusion® HF Reaction Buffer (Part No.: B0518S, New England Biolabs) OneTaq® Standard Reaction Buffer (Part No.: M0480, New England Biolabs)
Polymerases	2 U/μl Phusion® High-Fidelity DNA Polymerase (Part No.: M0530S, New England Biolabs) 5 U/μl Ampliqon TEMPase Hotstart DNA Polymerase (Part No.: 220308, Biomol) *Taq* DNA Polymerase 5 U/μl (Part No.: D6677, Sigma-Aldrich) OneTaq® DNA Polymerase 5 U/μl (Part No.: M0480, New England Biolabs)
Potassium chloride	KCl, >99,5%, p.a., ACS, ISO (Part No.: 6781.1, Carl Roth GmbH)
Potassium dihydrogen phosphate	KH_2PO_4, >99%, p.a., ACS (Part No.: 3904.1, Carl Roth GmbH)
Primer	desalted; dissolved in ncf-water, 100 μM (Sigma-Aldrich)
Proteinase K	Proteinase, Tritirachium album serine (Part No.: A3830,Applichem)

Chemicals and enzymes	Name (Provider)
Restriction enzymes	all New England Biolabs *Taq*I, recombinant, 20 and 100 U/μl (Part No.: R0149 L) *Bam*HI, recombinant, 20 U/μl (Part No.: R0136 L) *Pvu*II, recombinant, 10 U/μl (Part No.: R0151 L) *Rsa*I, 10 U/μl (Part No.: R0167 L) *Eco*RI, 20 U/μl (Part No.: R0101 L) *Eco*RV, 20 U/μl (Part No.: R0195 L) *Nla*III, 10 U/μl (Part No.: R0125 S)
Sodium chloride	Sodium chloride (Part No.: S5886-1KG, Sigma-Aldrich)
Sodium dodecyl sulfate (SDS)	Ultra pure (Part No.: 2326.1, Carl Roth)
Tissue freezing medium	Jung (Part No.: 020108926, Leica Microsystems)
Tris	Tris(hydroxymethyl)aminomethane (Part No.: 154563-1KG, Sigma-Aldrich)
Uracil-DNA Glycosylase (UDG)	5 U/μl (Part No.: M0280S, New England Biolabs)

5.1.4 Buffers and solutions

Buffers and solutions	Preparation/Provider
0.5 M EDTA	186.12 g EDTA add to 800 ml distilled water, stir vigorously and adjust the pH to 8.0 with NaOH; adjust volume to 1 l with distilled water
10x PBS-CMF	80 g NaCl (1.37 M), 2 g KCl (27 mM), 11.1 g Na_2HPO_4 (78 mM), 2 g KH_2PO_4 (15 mM)
1x TAE-buffer	Adjust 200 ml 50x TAE-buffer to 10 l with distilled water
50x TAE-buffer	242 g Tris are dissolved in 800 ml distilled water and 57.1 ml glacial acetic acid and 100 ml 0.5 M EDTA (pH 8.0) are added. Volume is adjusted to 1 l
6 M NaCl-Solution	350.4 g NaCl add to 1 l distilled water
Agarose gel preparation	1.5% (w/w); 1.5 g agarose powder are added in a 300 ml Erlenmeyer flask and made up to 100 g with 1xTAE-buffer 3.0% (w/w); 3.0 g agarose powder are added in a 300 ml Erlenmeyer flask and made up to 100 g with 1xTAE-buffer
Bovine serum albumin (BSA)	10 mg/ml purified BSA 100x (Part No.: B9001S, New England Biolabs)
CD-fetale calf serum (FCS)	charcoal/dextran treated FCS (Part No.SH30068.03, Hyclone, South Logan, USA)
Cell culture media	Minimum Essential Medium Eagle (MEM), with Earle's salts and non-essential amino acids, without L-glutamine, phenol red and sodium bicarbonate, Modified, powder (Part No.: M3024, Sigma-Aldrich)

Buffers and solutions	Preparation/Provider
	Nutrient Mixture F-10 Ham (F-10), with sodium bicarbonate, without L-glutamine, liquid, sterile-filtered (Part No.: N6013, Sigma-Aldrich)
	Nutrient Mixture F-12 Ham (F-12), with L-glutamine, without sodium bicarbonate, powder (Part No.: N6760, Sigma-Aldrich)
	L-15 Medium (Leibovitz), without L-glutamine, liquid, sterile-filtered (Part No.: L5520, Sigma-Aldrich)
	Dulbecco's Modified Eagle's Medium (DMEM) - high glucose, with 4500 mg/l glucose and L-glutamine, without sodium bicarbonate, powder (Part No.: D5648, Sigma-Aldrich)
	DMEM - low glucose, with 1000 mg/l glucose and L-glutamine, without sodium bicarbonate and phenol red, powder (Part No.: D2902, Sigma-Aldrich)
	DMEM/F-12, with 15 mM HEPES and sodium bicarbonate, without L-glutamine and phenol red, liquid, sterile-filtered (Part No.: D6434, Sigma-Aldrich)
	DMEM, HEPES modification, with 4500 mg/l glucose, 25 mM HEPES, and sodium bicarbonate, without L-glutamine and sodium pyruvate, liquid, sterile-filtered (Part No.: D6171, Sigma-Aldrich)
Cresyl violet	cresyl violet acetate (Sigma-Aldrich) is dissolved at a concentration of 1% (w/v) in 50% EtOH at room temperature and stirred overnight. Next day, the staining solution is filtered before use to remove unsolubilized powder.
dNTP mix	dNTP/dUTP-Mix: 10 μl 100 mM dATP + 10 μl 100 mM dCTP + 10 μl 100 mM dGTP + 20 μl 100 mM dUTP + 60 μl ncf-water (ACGU 10 mM/10 mM/10 mM/20 mM)
Eosin	Esoin Y (Part No.: E4009-5G, Sigma-Aldrich)
FCS	FCS (Part No.: 10270, LOT:41A1692K, Invitrogen, Karlsruhe, Germany)
FCS, heat inactivated	incubate FCS (Part No.: 10270, LOT:41A1692K, Invitrogen, Karlsruhe, Germany) at 56°C (water bath) for 30 min
Hematoxylin	Mayer's Hematoxylin (LOT No. 11639 , Leica Microsystems)
Krebs-Henseleit buffer	D-glucose (2 g/l), $MgSO_4$ (anhydrous, 0.141 g/l), KH_2PO_4 (0.16 g/l), KCL (0.35 g/l), NaCl (6.9 g/l), $CaCl_2x2H_2O$ (0.373 g/l), $NaHCO_3$ (2.1 g/l), dissolve salts in the given order, with 1 M HCl adjust pH to 7.2, sterile filtration
L-glutamine	200 mM, (Part No.: G7513, Sigma-Aldrich)
MEM amino acids	MEM Amino Acids (50x) solution, without L-glutamine, liquid, sterile-filtered (Part No.: M5550, Sigma-Aldrich)

Buffers and solutions	Preparation/Provider
MEM vitamins	MEM Vitamin Solution (100x), sterile-filtered (Part No.: M6895, Sigma-Aldrich)
NEB buffer 4	10x NEB Buffer 4 (Part No.: B7004S, New England Biolabs)
PBS-CMF	dilute 10x PBS-CMF 1:10, adjust pH to 7.4
Penicillin-Streptomycin	5000 U/ml and 5 mg/ml, respectively (Part No.: P4458, Sigma-Aldrich)
Primer mix	Add 10 μl of 100 μM forward and reverse primer to 180 μl ncf-water for an end concentration of 5μM of each primer.
SDS 10% (w/w)	Dissolve 100 g SDS in 900 ml distilled water (stir vigorously at 68°C) and adjust pH to 7.2; adjust volume to 1 l with distilled water
SE-buffer	Add 4.38 g NaCl, 50 ml 0.5 M EDTA and 100 ml 10% (w/w) SDS to 800 ml distilled water and adjust pH to 8.0 with NaOH; adjust volume to 1 l with distilled water
Sodium pyruvate	Sodium pyruvate solution (Part No.: S8636, Sigma-Aldrich)
SYBR Green	SYBR® Green I nucleic acid gel stain 10.000x concentrated in DMSO; 500 μl (Part No.: S-7563, Molecular Probes, Darmstadt, Germany)
SYBR Green Dilution I	Dilute 1 μl SYBR Green with 99 μl 1x TAE-buffer
SYBR Green Dilution II	Dilute 10 μl SYBR Green Dilution I with 990 μl 1x TAE-buffer
SYBR Green Dilution II (electrophoresis)	Dilute 10 μl SYBR Green Dilution I with 990 μl 50% (v/v) glycerin
SYBR Green Dilution III	Dilute 10 μl SYBR Green Dilution II with 990 μl 1x TAE-buffer
SYBR Green Dilution IV	Dilute 10 μl SYBR Green Dilution II with 990 μl 1x TAE-buffer
TE-buffer	Add 5 ml 1 M Tris and 0.1 ml 0.5 M EDTA to 400 ml distilled water and adjust pH to 8.0 with HCl; adjust volume to 1 l.
Trypsin	sterilized using filtration (Part No.: T4549, Sigma-Aldrich)

5.1.5 Commercial kits and reagents

Kit/reagent	Name, properties, provider
Magnetic beads	Dynabeads® Kilo Base Binder Kit™, Invitrogen, Germany
Deoxyribonuclease	Deoxyribonuclease 1 Amplification Grade, AMPD1-1KT, Sigma-Aldrich
Reverse Transcription Kit	High Capacity cDNA Reverse Transcription Kit, 200 reactions, Part No.: 4368814, Life Technologies
PCR product and Gel Band Purification	Illustra GFX PCR DNA and Gel Band Purification Kit, Part No.: 28-9034-70, GE Healthcare, Solingen, Germany
RNA isolation	GenElute™ Mammalian Total RNA Miniprep Kit (Part No.: RTN70, Sigma-Aldrich) RNAqueous R-Micro (Part No.: AM1931, Life Technologies)
Human Mammary Gland RNA*	Human Mammary Gland Total RNA*, Catalog No.: 636576, Lot No.: 5090034, concentration 1 μg/μl in DEPC-treated water, Clonetch Laboratories INC., Mountain View, California, USA.
PCR master mixes	TaqMan® SNP Genotyping Assay, 20x, Part No.: 4362691, Life Technologies (Assay no: C__25746809_50 and C___2091255_30) TaqMan® Gene Expression Assays, 20x, Part No.: 4448892, Life Technologies (Assays: Tables 4, 5, 6) TaqMan® Gene Expression Master Mix, 2x, Part No.: 4369016, Life Technologies TaqMan® GTXpress™ Master Mix, 2x, Part No.: 4403311, 1 ml, Life Technologies TaqMan® PreAmp Master Mix, 2x, Part No.: 4384266, 1 ml, Life Technologies
Preamplification pool	TaqMan® Custom PreAmp pool, 2x, Part No.: 4441856 , 1 ml, Life Technologies

*Source: normal human mammary gland pooled from a 27 year old Caucasian female; cause of death: sudden death.

5.1.6 Customized TaqMan® low density arrays

The TaqMan® low density array (TLDA) contains eight sample-loading lines each connected by a micro channel to 48 miniature reaction chambers for a total of 384 wells per card. Gene specific exon-spanning primers and TaqMan® probes are factory-designed and embedded in each well. We choose up to 47 genes from Applied Biosystems Assays-on-Demand ™ Gene Expression Products. For determination of gene expression three customized TLDAs are used with different layouts:

1. WNT-screen (human), 16 transcripts in triplicate, eight samples (Part No.: 4346798, Life Technologies)

2. WNT32 (human), 32 transcripts in triplicate (transcripts 1-16 in slot number 1, 3, 5, 7 and transcripts 17-32 in slot number 2, 4, 6, 8), four samples (Part No.: 4346799, Life Technologies)

3. Polymorphism I+II (human), 48 transcripts in single reaction, eight samples (Part No.: 4342253, Life Technologies)

Table 4: Transcripts with abbreviation and identification number of TaqMan® Gene Expression Assay (assay number) of customized TLDA-WNT-screen designed for the relative quantification of these 16 transcripts including internal card control 18S.

Transcripts	***Abbreviation***	***Assay number***
Actin, β	ACTβ	Hs99999903_m1
Androgen receptor	AR	Hs00907244_m1
Estrogen receptor 1 (ERα)	ESR1	Hs00174860_m1
Estrogen receptor 2 (ERβ)	ESR2	Hs00230957_m1
Eukaryotic 18S rRNA	18 S	Hs99999901_s1
Glycogen synthase kinase 3, β	GSK3 β	Hs00275656_m1
Homeobox A10	Hoxa 10	Hs00172012_m1
Hypoxanthine-guanine phosphoribosyltransferase 1	HPRT1	Hs99999909_m1
Alkaline phosphatase, placental	ALPP	Hs01654626_s1
Progesterone receptor	PGR	Hs00172183_m1
Protein kinase C, β	PKC β	Hs00176998_m1
Protein kinase C, δ	PKC δ	Hs00178914_m1
Secreted frizzledrelated protein 4	SFRP4	Hs00180066_m1
Winglesstype integration site family, member 5a	WNT5A	Hs00180103_m1
Winglesstype integration site family, member 7a	WNT7A	Hs00171699_m1
βTransducin repeat containing	BTRC	Hs00182707_m1

Table 5: Transcripts with abbreviation and identification number of TaqMan® Gene Expression Assay (assay number) of customized TLDA-WNT32 designed for the relative quantification of these 32 transcripts including internal card control 18S.

Transcripts	***Abbreviation***	***Assay number***
Hypoxanthine phosphoribosyltransferase	HPRT	Hs99999909_m1
Actin, β	ACTβ	Hs99999903_m1
Wingless-type MMTV integration site family, member 7A	WNT7A	Hs00171699_m1
Eukaryotic 18S rRNA	18S	Hs99999901_s1
Wingless-type MMTV integration site family, member 5A	WNT5A	Hs00180103_m1
Estrogen receptor 1 (ERα)	ESR1	Hs00174860_m1
Estrogen receptor 2 (ERβ)	ESR2	Hs00230957_m1
Alkaline phosphatase, placental	ALPP	Hs01654626_s1
Progesterone receptor	PGR	Hs00172183_m1
Axin 2	AXIN2	Hs00610344_m1
Frizzled homolog 7 (Drosophila)	FZD7	Hs00275833_s1
Msh homeobox 1	MSX1	Hs00427183_m1
Beta-transducin repeat containing	BTRC	Hs00182707_m1
Homeobox A10	HOXA10	Hs00172012_m1
Homeobox A11	HOXA11	Hs00194149_m1
Calcium/calmodulin-dependent protein kinase II beta	CAMK2B	Hs00176186_m1
Tumor necrosis factor	TNF	Hs00174128_m1
Matrix metallopeptidase 1 (interstitial collagenase)	MMP1	Hs00233958_m1
Protocadherin 8	PCDH8	Hs00159910_m1
Lymphoid enhancer-binding factor 1	LEF1	Hs00212390_m1
Transcription factor 7 (T-cell specific, HMG-box)	TCF7	Hs00175273_m1
Transcription factor 7-like 1 (T-cell specific, HMG-box)	TCF7L1	Hs00229841_m1
Transcription factor 7-like 2 (T-cell specific, HMG-box)	TCF7L2	Hs00181036_m1
Wingless-type MMTV integration site family, member 7B	WNT7B	Hs00536497_m1
Frizzled homolog 10 (Drosophila)	FZD10	Hs00273077_s1
Frizzled homolog 5 (Drosophila)	FZD5	Hs00361869_g1
Secreted frizzled-related protein 4	SFRP4	Hs00180066_m1
Dickkopf homolog 2 (*Xenopus laevis*)	DKK2	Hs00205294_m1
Frizzled homolog 4 (Drosophila)	FZD4	Hs00201853_m1
Low density lipoprotein receptor-related protein 5	LRP5	Hs00182031_m1
Low density lipoprotein receptor-related protein 6	LRP6	Hs00233935_m1
Protein kinase C, β	PRKCB1	Hs00176998_m1

Table 6: Transcripts with abbreviation and identification number of TaqMan® Gene Expression Assay (assay number) of customized TLDA designed for the relative quantification of these 52 transcripts, including internal card control 18S. *, only on TLDA-Polymorphisms I. ', only on TLDA-Polymorphism II. °, determined using TaqMan® probe-based real time PCR in 96 well-format.

Transcripts	***Abbreviation***	***Assay number***
Apoptosis/cell cycle		
BCL2 Antagonist of cell death	BAD	Hs00188930_m1
BCL2 modifying factor	BMF	Hs00372937_m1
BCL2-associated X protein	BAX	Hs99999001_m1
Cyclin D1	CCND1	Hs99999004_m1
Cyclin-dependent kinase inhibitor 1 A	CDKN1A	Hs01121172_m1
Cyclin-dependent kinase inhibitor 1 B	CDKN1B	Hs00153277_m1
GATA binding protein 3	GATA3	Hs00231122_m1
Growth arrest and DNA-damage-inducible, protein 45 alpha	GADD45A	Hs00169255_m1
Murine double minute 2	MDM2	Hs01066930_m1
Proliferation-related Ki-67 antigen	MKI67	Hs00267195_m1
Stratifin (14-3-3 Sigma)	SFN	Hs00356613_m1
Trefoil factor 1 (pS2, Breast cancer estrogen-inducible protein)	TFF1	Hs00907239_m1
Tumor protein p53	TP53	Hs00153340_m1
E2 metabolism		
17beta-Hydroxysteroid dehydrogenase 1	HSD17B1	Hs00166219_g1
17beta-Hydroxysteroid dehydrogenase 2	HSD17B2	Hs00157993_m1
Catechol-O-methyltransferase	COMT	Hs00241349_m1
Cytochrome P450, family 1, subfamily a, polypeptide 1	CYP1A1	Hs00153120_m1
Cytochrome P450, family 1, subfamily a, polypeptide 2*	CYP1A2*	Hs00167927_m1*
Cytochrome P450, family 1, subfamily b, polypeptide 1	CYP1A1	Hs00164383_m1
Cytochrome P450, family 19, subfamily a, polypeptide 1 (Aromatase)	CYP19A1	Hs00903411_m1
Glutamate-cysteine ligase, catalytic subunit	GCLC	Hs00155249_m1
Glutathione S-transferase mu 1*	GSTM1*	Hs02341469_m1*
Glutathione S-transferase pi 1	GSTP1	Hs00168310_m1
Glutathione S-transferase theta 2	GSTT1	Hs00184475_m1
NAD(P)H dehydrogenase, quinone 1	NQO1	Hs01045994_m1
Steroid-sulfatase	STS	Hs00996676_m1
Sulfotransferase family 1E, estrogen-preferring, member 1	SULT1E1	Hs00193690_m1
Sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1	SULTA1	Hs00738644_m1
Sulfotransferase family, cytosolic, 1A, phenol-preferring, member 2	SULT1A2	Hs02340929_g1
Sulfotransferase family, cytosolic, 1A, phenol-preferring, member 3	SULT1A3;SULT1A4	Hs00413970_m1
Sulfotransferase family, cytosolic, 2A, member 1	SULT2A1	Hs00234219_m1
UDP glycosyltransferase 1 family, polypeptide A4	UGT1A4;UGT1A3	Hs01592480_m1
UDP glycosyltransferase 1 family, polypeptide A8	UGT1A8	Hs01592482_m1
UDP glycosyltransferase 2 family, polypeptide B7	UGT2B7	Hs00426592_m1
UDP glycosyltransferase 1 family, polypeptide A1°	UGT1A1°	Hs02511055_s1°
UDP glycosyltransferase 1 family, polypeptide A10°	UGT1A10 °	Hs02516990_s1°
Housekeeping gene		
Glucuronidase, beta'	GUSB'	Hs99999908_m1'
Hypoxanthine-guanine phosphoribosyltransferase 1	HPRT1	Hs99999909_m1
TATA box binding protein'	TBP'	Hs00427620_m1'
Mandatory endogenous control		
Eukaryotic 18S rRNA	18S	Hs99999901_s1
Nuclear receptors in E2 metabolism		
Aryl hydrocarbon receptor	AHR	Hs00169233_m1
Aryl hydrocarbon receptor nuclear translocator	ARNT	Hs01121918_m1
Estrogen receptor alpha	ESR1	Hs00174860_m1
Estrogen receptor beta	ESR2	Hs00230957_m1
Nuclear factor (erythroid-derived 2)-like 2 (NRF2)	NFE2L2	Hs00975960_m1
Nuclear receptor, sub family 1, group I, member 2	NR1I2 (PXR)	Hs00243666_m1
Nuclear receptor, sub family 1, group I, member 3	NR1I3 (CAR)	Hs00901571_m1
Progesterone receptor	PGR	Hs00172183_m1
Paracrine signal transduction		
Amphiregulin (SDGF, Schwannoma derived growth factor)	AREG	Hs00950669_m1
Transforming growth factor beta 1	TGFB1	Hs00171257_m1
Wingless-related MMTV integration site 4	WNT4	Hs00229142_m1
Wingless-related MMTV integration site 5A	WNT5A	Hs00180103_m1

5.1.7 Software

Software	Version, Provider
ChemBioDraw Ultra	Version 12.0, CambridgeSoft, Cambridge, USA
ExpressionSuite	Version 1.0.4, Life Technologies
Quantity One®	Version 4.6.8, Bio-Rad Laboratories GmbH
i-controle	Version 1.10, Tecan
Leica Laser Microdissection	Version 7.3.1.4552, Leica Microsystems
LASAFExpressView	Leica Microsystems
OriginPro	Version 8.6 G, OriginLab Corporation, Northampton, MA 01060, USA
Phyton	Version 2.5, Open source, URL https://www.python.org/download/ releases/2.5/
R	Version 3.0.1, R Core Team (2013), R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/
SDS-sequence detection system	Version 2.4, Life Technologies
Strawberry Perl	5.18.2.1-64bit, Open source, URL http://strawberryperl.com/ releases.html
YANAsquare	Version 21.03.2012, Julius-Maximilians-Universität Würzburg: YANAsquare. URL: http://www.biozentrum.uni-wuerzburg.de/yana.html

5.2 Methods

5.2.1 Cell culture

Cells are seeded in cell culture flasks (175 cm^2, Greiner bio-one; 1.7x10^4 cells per cm^2). All cell lines are grown at 37°C and 100% humidity and 5% CO_2, except MDA cells are grown without exposure to CO_2. All cell lines are cultured until 80-90% confluence, trypsinized, pelleted and stored at -80°C until isolation of DNA. All media are supplemented with 100 U/ml penicillin and 100 μg/ml streptomycin.

MCF-7 cells

MCF-7 BUS cells (Soto et al., 1995), kindly provided by Ana Soto (Tufts University, Boston, MA, USA), are grown in DMEM high glucose medium (normal culture media) supplemented with 1 mM pyruvate, 2.25 g/l sodium bicarbonate and 5% heat-inactivated fetal calf serum (FCS, Invitrogen™ Life Technologies). For steroid-free cultivation, MCF-7 cells are seeded in DMEM high glucose medium (10,000 cells/cm^2 for 24 h) and then the medium is changed to in phenol red-free DMEM medium (Sigma-Aldrich) supplemented with 5% charcoal/dextran-treated FCS (Hyclone), 4.5 g/l D(+)glucose, 15 mM HEPES buffer and 2.25 g sodium bicarbonate (steroid-free culture media). Cells are trypsinated after 48 h.

Ishikawa cells

Ishikawa cells (Nishida, 2002), kindly provided by Ken Korach (National Institute of Environmental Health Sciences, NC, USA), are cultured in phenol red-free DMEM:F-12 (Sigma-Aldrich) supplemented with 2.5 mM L-glutamine and 10% FCS.

Hela S3 cells

Hela S3 cells are grown in F12-Medium (Sigma-Aldrich) supplemented with 10% FCS and 2 mM L-glutamine.

MDA cells

MDA-MB-453 cells (DSMZ, German Collection of Microorganisms and Cell Cultures) are grown in Leibovitz L-15 media (Sigma-Aldrich) supplemented with 10% FCS. The cells are cultured without exposure to 5% CO_2.

HT-29 cells

HT-29 cells, obtained from the German Collection of Microorganism and Cell Cultures (Braunschweig, Germany) are grown in DMEM (Sigma-Aldrich) supplemented with 10% FCS.

HCV cells

Human chorion villi cells, self isolated, are grown in F10-Medium (Sigma-Aldrich) supplemented with 15% FCS and 1 mM L-glutamine.

AG01522C cells

AG0 cells are grown in MEM (Sigma-Aldrich) supplemented with 16% FCS, MEM vitamins (Sigma-Aldrich), MEM amino acids (Sigma-Aldrich), 2.2 g/l $NaHCO_3$ and 2 mM L-glutamine.

5.2.2 Preparation of fresh mammary gland tissue

Normal human mammary gland tissues are obtained from women who underwent a mamma reduction surgery for cosmetic reasons. The female donors answered a questionnaire concerning their age, height and weight (or BMI), number of pregnancies, intake of hormone active drugs (HAD), smoking habit and alcohol consumption. The study has been approved by the Ethics Committee of the Faculty of Medicine at the University of Würzburg and written consent was obtained from each healthy volunteer. After removal, tissue is put in ice

cold Krebs-Henseleit buffer, separated from surrounding fatty tissue, portioned, flash frozen in liquid nitrogen (all steps within 10-15 minutes after removal) and stored at -80°C.

5.2.3 Grinding of mammary gland tissue

For better homogenization the mammary gland tissue is finely ground in a special precision mortar (self-construction by internal facility) by hammering in liquid nitrogen. Approx. 200 mg of mammary gland tissue are hammered with 4 times of 6 strokes. After each set of strokes the powder is scraped off the mortar bottom. The fine powder is then filled back in a cryo-tube and stored at $-80\,^{\circ}\mathrm{C}$ until RNA isolation. During the whole procedure the equipment and the tissues are kept in liquid nitrogen.

5.2.4 Cryostat sectioning

First, the specimen (tissue aliquot) is attached to the specimen disk by embedding the tissue in tissue freezing medium (Jung Tissue freezing medium™, Leica): The pre-cooled specimen disk is covered with the viscous freezing medium and the specimen is then placed on top of the freezing medium. The specimen disk is attached to the activated Peltier element on the freeze shelf in the cryostat for fast freezing. After complete freezing of the freezing medium, the specimen disk is inserted into the specimen head. The tissue is then cut into 8 μm sections. The chamber temperature is set to $-20\,^{\circ}\mathrm{C}$ and object temperature (specimen head) is set to $-25\,^{\circ}\mathrm{C}$. For sectioning disposable blades are used. UVC disinfection for at least ten min between cutting of two different samples. The section is then unclamped by thawing it to the object slide (membrane slide, Leica Microsystems).

5.2.4.1 Section staining

The short staining procedure of the section located on the slide using Cresyl Violet (Sigma-Aldrich) colors the nuclei violet and the cytoplasm weak violet. The slides with tissue section are stained with short protocol avoiding RNA degradation caused by RNases. Endogenous RNases in frozen sections may still be active after the short fixation step. Therefore all incubation steps are kept as short as possible. RNase-free water (Thermo Scientific) and ethanol (Applichem) are used for all steps.

Procedure:

1. fixation (2 min, 70% ethanol)
2. stain 30 sec into 1% Cresyl Violet acetate solution
3. remove excess stain on absorbent surface

4. dip into 70% ethanol
5. dip into 100% ethanol
6. air-dry shortly (1-2 min)

To determine the lobule type of the sample sections are stained with Cresyl Violet (described above) or Hematoxylin (Leica) and Eosin (Sigma-Aldrich), which colors the nuclei blue and the cytoplasm pink.

Procedure:

1. fixation (2 min, 70% ethanol)
2. rinse in distilled water
3. stain 2 min in Hematoxylin solution (Leica)
4. remove excess stain on absorbent surface
5. rinse in tap water (1 min)
6. stain 10 sec in Eosin Y solution
7. dip into 70% ethanol
8. dip into 100% ethanol
9. air-dry shortly (1-2 min)

5.2.5 Laser capture microdissection

For laser capture microdissection frozen aliquots of human mammary gland tissue are cut into 15 μm sections with a cryostat (Chapter 5.2.4), placed on a membrane slide and the sections are stained with Cresyl Violet (Chapter 5.2.4.1). The laser of the Laser Microdissection System (Leica) is calibrated according to manufacturers instructions. Epithelial cells are cut out using 40x objective (Leica) and the following settings of the laser:

Power	38
Aperture	11
Speed	7
Specimen balance	4
Offset	180

Every isolated element is stored in a shape list including area of the element. A micrograph is taken before and after laser cutting, every micrograph is automatically stored in the database (LASAFEXExpressView, Leica). The elements fall into a dry tube (0.2 ml) cap located under the slide holder.

5.2.6 RNA isolation

Depending on material (tissue homogenate, cultured cells and isolated cells of laser capture microdissection) different commercial available kits for isolation of RNA are used:

Homogenate and cultured cells

RNA of cultured cells and mammary gland tissue is isolated using the GenElute™ Mammalian Total RNA Miniprep Kit (Sigma-Aldrich). Since mammary gland is a fibrous tissue with integrated adipocytes a few additional steps are included in RNA isolation procedure to improve RNA yield. First, the mammary gland tissue is powdered as described in Chapter 5.2.3 and homogenated in lysis buffer according to manufacturer's protocol followed by a 10 min proteinase K digestion at 55°. Then, the fat is extracted using chloroform. The aqueous phase is collected and RNA isolated according to manufacturer's protocol and eluted in 50 μl ncf-water.

DNase digestion:

The RNA solution is mixed with 5 μl of reaction buffer and 5 μl DNase solution. The solution is allowed to stand 15 min at room temperature. Then, 5 μl of stop solution is added, mixed and incubated at 70 °C for 10 min in a water bath. A reagent blank is prepared at the same time. The RNA solution is placed on ice and the RNA concentration is determined (Chapter 5.2.8). Until further use the RNA is stored at −80 °C.

Isolated cells of laser capture microdissection

For RNA isolation of isolated cells of laser capture microdissection the RNAqueous Micro Kit (AM1931, Ambion) is used. All steps are carried out according to manufacture's protocol. This kit also includes digestion of the remaining DNA with DNase. DNase digestion is also carried out according to manufacture's protocol.

5.2.7 Isolation of total DNA

Trypsinated cells are taken up in 10 ml of medium and centrifuged at 260 x g for 5 min. The pellet is taken up in 5 ml SE-buffer, which contains 200 μg/ml proteinase K. The mixture is incubated at 55°C in a shaking water bath overnight. After incubation, 2 ml of NaCl solution (6 M) are added and mixed well. Then, 7 ml chloroform are added. Agitate well at room temperature for 30 seconds. Phases are separated by centrifuging (4°C, 1800xg) for 5 min. The upper, aqueous phase is removed, transferred to a new centrifuge tube and carefully mixed with an equal volume of isopropanol. Then, the fibrous DNA precipitate. The mixture is centrifuged for 5 min (4°C, 1800xg) and the DNA is transferred with a plastic pasteur pipette into a 1.5 ml centrifuge tube. 1 ml of 70% ethanol is added and stored for 0.5 h in refrigerator. Then, it is again centrifuged (4°C, 1800xg). The ethanol is decanted

and the remaining ethanol is carefully evaporated in a stream of N_2 or air dried. The DNA is dissolved at room temperature overnight in 200 μl of TE-buffer (pH 8.0). The next day, the DNA concentration is quantified as described in Chapter 5.2.8. For human mammary gland tissue, 2 g tissue are grounded (Chapter 5.2.3) and suspended 20 ml SE-buffer and further treated as described above. The amounts of all reagents are increased according to 4-fold amount of SE-buffer.

5.2.8 RNA/DNA quantification

The concentration of RNA or double stranded DNA is determined using NanoQuant Plate™ for TECAN plate reader (Tecan). For this purpose, 2.0 μl of reagent blank of the DNase digestion (RNA isolation) or TE-buffer (DNA isolation) are used for individual blanking, meaning each blank value is thereby assigned to each used well, and samples are blanked with the appropriate, well-specific blank value. Then, 2.0 μl of the RNA/DNA solution are pipetted to the respective wells (blanked wells) on the NanoQuant Plate™ and the absorbance is measured at 260 nm and 280 nm. To evaluate the purity of RNA and DNA samples, the ratio of the absorbance at 260 nm/280 nm is calculated. Good quality DNA have a ratio between 1.8 and 2.0 and good quality RNA between 1.9 and 2.1.

$$C = \frac{A}{(e * l)} \tag{1}$$

C: Concentration of the nucleic acid in μg/ml.

A: Absorbance at 260 nm.

l: Width of the cuvette (0.2 cm for the NanoQuant Plate™)

e: Extinction coefficient of double standed DNA (50 μg/ml) or RNA (40 μg/ml)

For determination of DNA concentrations smaller than 40 ng/μl (DNA-standard, Chapter 5.2.12.7) the photometric analysis is carried out in a photometer (Shimadzu GmbH) using 50 μl-cuvettes (UVette®, Eppendorf), as blank value the solvent is used in the same cuvette. The DNA concentration of sample is calculated using the absorbance at 260 nm.

The copy number of DNA is calculated as follows:

The DNA concentration of the DNA standard (pg/μl) is converted in copies of dsDNA per μl (copies/μl):

$$\text{Number of copies} = \frac{\text{A} * \text{NA}}{(B * M) * 1x10^9 \text{ ng/g}} \tag{2}$$

A= Amount of DNA (ng)

B= length of dsDNA (bp)

M= Average mass of 1 bp dsDNA

NA= Avogadro constant = $6.0221 \text{x} 10^{23}$ molecules/mole

5.2.9 Reverse transcription

RNA is reversely transcribed into cDNA using the High Capacity Reverse Transcription Kit (Life Technologies). Whenever possible 1 μg RNA is reversely transcribed in 20 μl stock according to Table 7, but maximum 8 μl of DNase digested RNA solution (Chapter 5.2.6) is used. The temperature program (Table 8) is performed according to manufacture's protocol.

Table 7: Pipetting scheme of reverse transcription using High Capacity Reverse Transcription Kit (Life Technologies). Final volume 20 μl. Template: RNA (Chapter 5.2.6).

Component	Concentration	Volume (μl)
RT buffer	10x	2
RT Random Primers	10x	2
dNTP Mix	25x (100 mM)	0.8
MultiScribe® Reverse Transcriptase	50 U/μL	1
Template	-	up to 8
Ncf-water	-	variable

Table 8: Temperature program for reverse transcription using High Capacity Reverse Transcription Kit (Life Technologies).

Temperature (°C)	Time
25	10 min
37	120 min
85	5 min
4	∞

5.2.10 Transcript level quantification

Transcript levels of all target genes (Tables 4, 5 and 6) are quantified using commercial available TaqMan® Gene Expression Assays (96-well format) or TLDA. Both contain specific forward and reverse primers, as well as the TaqMan® probes labeled with FAM™ dye for the determination of the target sequences. To quantify low expressed transcripts a preamplification reaction prior real time PCR is used. Additionally, the HPRT-TaqMan® Gene Expression Assays (96-well format) is used to verify successful reverse transcription and to calculate the amount of cDNA required for preamplification.

5.2.10.1 TaqMan® Gene Expression Assays

To verify successful reverse transcription and to calculate the amount of cDNA required for preamplification, a 20 μl PCR with HPRT-TaqMan® Gene Expression Assay primers is performed with a 7900HT fast real-time PCR system. Likewise, the transcript levels of GSTP1 and NQO1 (in cDNA), UGT1A1 and UGT1A10 (in preamplified cDNA) are quantified using 20 μl PCR with TaqMan® Gene Expression Assay primers. The Ct values are automatically calculated using SDS software 2.4 (Chapter 5.1.7). The initial copy number for each transcript is calculated based on 1x10^{12} copies at Ct and an efficiency of 100% (as specified by the manufacturer) according to formula: $n0 = \frac{n_{Ct}}{E^{Ct}} = \frac{1x10^{12}}{2^{Ct}}$.

Table 9: Pipetting scheme of TaqMan® probe-based real time PCR using TaqMan® Gene Expression Assay primers (Life Technologies). Final volume 20 μl. Template: cDNA, preamplified cDNA or pre-preamplified cDNA, if necessary diluted in ncf-water.

Component	Concentration	Volume (μl)
Gene expression master mix	2x	10
Gene expression assay	20x	1
Template	-	1
Ncf-water	-	8

Table 10: Temperature program for TaqMan® probe-based real time PCR using TaqMan® Gene Expression Assay primers (Life Technologies).

Repetition	Step	Temperature (°C)	Time
1x	Denaturation	95	2 min
40x	Denaturation	95	30 sec
	Annealing	60	30 sec

5.2.10.2 Preamplification and pre-preamplification

Without preamplification, Ct values of several target transcripts are too high for quantification (i.e. Ct value >33). The preamplified cDNA is adjusted to reach a Ct value of 17-18 for the HPRT transcript (Table 11). Preamplification reactions are run on a Primus 25 Thermal Cycler (MWG Biotech, Table 14). After preamplification, the Ct values of the HPRT transcript are determined again. The preamplification pool is either custom made by Life Technologies (for TLDA-PolymorphismII) or self pipetted: TaqMan® Gene Expression Assays for each investigated transcript are pooled and diluted with ncf-water to a final concentration of 0.2-fold each in the preamplification pool according to the manufacturers instructions.

Table 11: Pipetting scheme of preamplification (PreAmp, Life Technologies). Final volume 20 μl. Template: cDNA or preamplified cDNA, if necessary diluted in ncf-water.

Component	Concentration	Volume (μl)
TaqMan® PreAmp Master Mix	2x	10
PreAmp pool	0.2x	5
Template	-	2
Ncf-water	-	3

Table 12: Temperature program of preamplification of TaqMan® probe-based real time PCR using preamplification pool (Life Technologies).

Repetition	Step	Temperature (°C)	Time
1x	Denaturation	95	10 min
10-14x	Denaturation	95	15 sec
	Annealing	60	4 min

5.2.10.3 TLDA

Customized TLDA enabling the analysis of 16-48 transcripts and four-eight samples per card is used: preamplified cDNA is diluted with ncf-water to achieve a Ct value for HPRT of 15-17 and mixed with TaqMan® Gene Expression PCR Master Mix (Table 13), loaded and run on a 7900HT system. The Ct values are automatically calculated using SDS software 2.4 or ExpressionSuite software (Life Technologies, Chapter 5.1.7). The initial copy number for each transcript is calculated based on $1x10^{12}$ copies at Ct and an efficiency of 100% (as specified by the manufacturer) according to formula in Chapter 5.2.10.1 and standardized to n0 of HPRT-transcript.

Table 13: Pipetting scheme of TLDA (Life Technologies). Final volume is 100 μl loaded into one slot. Template (cDNA, preamplified cDNA or pre-preamplified cDNA).

Component	Concentration	Volume (μl)
TaqMan® Gene expression Master Mix	2x	50
Template	-	5-15
Ncf-water	-	variable

Table 14: Temperature program of preamplification of TaqMan® probe-based real time PCR using TLDA-PolymorphismII (Life Technologies).

Repetition	Step	Temperature (°C)	Time
1x	Denaturation	94.5	10 min
40x	Denaturation	97	30 sec
	Annealing	59.7	1 min

5.2.11 Determination of polymorphisms

In this work two types of polymorphisms (SNP and CNP) are determined: SNPs are determined using allelic discrimination assays commercially available (Chapter 5.2.11.1) and restriction fragment length (RFL) PCR (Chapter 5.2.11.2). The CNP in GSTT1 is determined using conventional PCR (Chapter 5.2.11.3). For both methods total DNA is isolated as described in Chapter 5.2.7.

5.2.11.1 Allelic discrimination

The allelic discrimination is used to determine the SNP of COMT Val108/158Met (rs4680, C_25746809_50 inventoried TaqMan® SNP Genotyping Assay, Life Technologies) and the SNP of NQO1 Pro187Ser (rs1800566, C_2091255_30 inventoried TaqMan® SNP Genotyping Assay, Life Technologies). The TaqMan® SNP Genotyping Assays contain the sequence-specific forward and reverse primers to amplify the polymorphic sequence and two TaqMan® probes. One probe is labeled with VIC® dye detecting the sequence of allel 1 and the second probe is labeled with FAM™ dye detecting the sequence of allel 2 . The DNA, isolated from cells or tissue (Chapter 5.2.7) is diluted in TE-buffer with a final concentration of 50 ng/μl. The reaction is carried out in 5 μl in 384-well plates on a 7900 HT Fast Real time PCR System (Life Technologies) and set up as described in Table 15. Both detectors, VIC® and FAM™, are used for each well. An allelic discrimination plate read document is set up and the pre-read of the plate is performed, according to manufacture's protocol. Then, a PCR run is performed (Table 16). After PCR run, the allelic discrimination post-read of the plate is carried out. The automatic allel calls from allelic discrimination plot are reviewed and converted into genotypes manually.

Table 15: Pipetting scheme of allelic discrimination using TaqMan® SNP Genotyping assays (Life Technologies). Final volume 5 μl. Template (DNA).

Component	Concentration	Volume (μl)
TaqMan® GTXpress™ master mix	2x	2.5
Genotyping assay	20x	0.25
Template	50 ng/μl	1
Ncf-water	-	1.25

Table 16: Temperature program of the PCR run of the allelic discrimination using TaqMan® SNP Genotyping assays (Life Technologies).

Repetition	Step	Temperature (°C)	Time
1x	Denaturation	95	10 min
40x	Denaturation	95	15 sec
	Annealing	60	60 sec

5.2.11.2 RFL PCR

The RFL PCR for the detection of COMT Val108/158Met SNP is carried according to Yim et al. (2001). The PCR is performed (Tables 17 and 18) using Primus 96 Thermal Cycler (MWG Biotech).

Sequence of primers (Yim et al., 2001):
Forward (FW) primer: 5'-TCGTGGACGCCGTGATTCAGG-3'
Reverse (REV) primer: 5'-AGGTCTGACAACGGGTCAGGC-3'

The PCR product (217 bp) is then digested (incubation at 37°C for 3 h) using the restriction enzyme *Nla*III (New England Biolabs, recognition sequence 5'-CATG-3' (Yim et al., 2001)). For the Val allele, fragments with the length of 114 bp, 83 bp and 20 bp and for the Met allele (two restriction sites), fragments with the length of 96 bp, 83 bp, 20 bp, 18 bp are generated (three restriction sites). The heterozygous variant generates all mentioned fragments (Figure 13).

Table 17: Pipetting scheme for RFL PCR. Final volume 25 μl.

Component	Final concentration	Volume (μl)
10x incubation mix	1x	2.5
$MgCl_2$	2.5 mM	2.5
dNTPs	400 μM	1
Taq-DNA-polymerase	1U	0.25
Primer mix	μM	4
Template	2 ng/μl	1
Ncf-water	-	13.75 μl

Table 18: Temperature program for RFL PCR.

Repetition	Step	Temperature (°C)	Time
1x	Denaturation	95	2 min
30x	Denaturation	95	30 sec
	Annealing	58	30 sec
	Extension	72	1 min
1x	Extension	72	5 min
	Cooling	10	∞

PCR product is sequenced by Sanger using LGC Genomics GmbH (Berlin, Germany) DNA sequencing service. PCR products are prepared according to LGC Genomics' DNA sequencing protocol.

5.2.11.3 Determination of CNP

The CNP of GSTT1 is determined as described in Buchard et al. (2007). The method is modified concerning the annealing/elongation step: It is reduced from 7 min to 4 min at 68 °C, since the GSTT1 genotyping is of interest and the method is described for the simultaneous determination of polymorphisms in GSTP1, GSTM1 and GSTT1.

Sequences of primers (Buchard et al., 2007):

GSTT1 gene:	FW primer 5'-TCTTTTGCATAGAGACCATGACCAG-3'
	REV primer: 5'-CTCCCTACTCCAGTAACTCCCGACT-3'
	product length: 969 bp
GSTT1 deletion:	FW primer: 5'-GAAGCCCAAGAATGGGTGTGTGTG-3'
	REV primer: 5'-TGTCCCCATGGCCTCCAACATT-3'
	product length: 3106 bp

Table 19: Pipetting scheme for RFL PCR. Final volume 25 μl.

Component	Final concentration	Volume (μl)
OneTaq® Standard Reaction Buffer	1x	5
dNTPs	400 μM	0.625
OneTaq® DNA Polymerase	1.25 U	0.25
Primer mix	300 nM (gene), 40 nM (deletion)	1.5 (gene) + 1.5 (deletion)
Template	2 ng/μl	1
Ncf-water	-	15.125 μl

Table 20: Temperature program of RFL PCR for determination of GSTT1 CNP.

Repetition	Step	Temperature (°C)	Time
1x	Denaturation	94	3 min
	Denaturation	94	30 sec
35x	Annealing	68	4 min
1x	Extension	68	10 min
	Cooling	10	∞

5.2.12 RMCA

The RMCA is divided into synthesis of the uracil containing biotinylated DNA-probe, hybridization, *Taq*I digestion, copy number quantification and mutant detection.

5.2.12.1 Synthesis of the uracil containing biotinylated DNA-probe

The uracil-containing, 5'-biotin-terminated DNA-probe complementary to the target sequence in intron 6 of p53 gene is synthesized via PCR. For the synthesis of the uracil containing biotinylated DNA-probe from human genomic DNA, a primer pair with a biotinylated reverse primer (primer forward 5'-CCT GCC CTC AAC AAG ATG T-3', primer reverse: 5'-Biotin-CAT CAT GGA GAT AAC ACA GGC CCA AG -3) and a dNTP mix containing dUTP instead of dTTP is used.

Table 21: Pipetting scheme of PCR for synthesis of uracil containing biotinylated probe. Final volume 25 μl.

Component	Final concentration	Volume (μl)
Reaction buffer	1x	2.5
$MgCl_2$	2.5 mM	2.5
dNTP/dUTP-mix (ACGU)	200 μM dATP/dCTP/dGTP 400 μM dUTP	0.5
DNA Polymerase	1.25 U	0.25
Primer mix	400 nM	2.0
Template	100 ng/μl	1
Ncf-water	-	16.25 μl

Table 22: Temperature program of PCR for synthesis of biotinylated probe.

Repetition	Step	Temperature (°C)	Time
1x	Incubation	37	10 min
1x	Denaturation	95	15 min
47x	Denaturation	95	30 sec
	Annealing	60	30 sec
	Extension	72	40 sec
1x	Extension	72	5 min
	Cooling	10	∞

PCR products are pre-stained with SYBR-Green and separated on an 1.5% agarose gel in 1x Tris-acetic acid-ethylenediaminetetraacetic acid (TAE-buffer). Electrophoresis is conducted at 5 V/cm for 1 h (Chapter 5.2.13). DNA is visualized by UV-excitation and the PCR product (927 bp) is cut off and DNA isolated using Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and resuspended in TE-buffer (Chapter 5.2.13.2).

5.2.12.2 DNA restriction digest of genomic DNA

Approx. 120 μg human DNA is digested for 16 h at 37°C with five restriction enzymes (*Pvu*II, *Eco*RI, *Eco*RV, *Bam*HI, *Rsa*I, New England Biolabs) that do not cut within the target sequence of the TP53 gene.

Table 23: Pipetting scheme for restriction digestion. Final volume 500 μl. Template: human DNA of cells (120 μg) or tissue (300 μg).

Component	Final concentration per 500 μl	Volume (μl)
*Pvu*II	100 U	10
*Rsa*I	100 U	10
*Bam*HI	200 U	10
*Eco*RI	200 U	10
*Eco*RV	200 U	10
NEB 4 buffer (10x)	1x	50
BSA (10 μg/μl)	5 μg	0.5
Template	120 μg (300 μg)	variable
Ncf-water	-	variable

5.2.12.3 Microcon purification of DNA restriction Digest

The restriction digest (Chapter 5.2.12.2) is purified using Microcon YM-50 according to manufacture's protocol: Centrifuge 10 min at 14,000xg and wash twice in 1 M NaCl by centrifuging 10 min at 14,000xg (discarding the flow through each time). The purified digest is eluted by inverting the filter and centrifuging at 1,000xg for 2 min. The flow through containing purified digest is adjusted to 500 μl with 1 M NaCl.

5.2.12.4 Hybridization

5 μl of Dynabeads® kilo base binder kit™ (Invitrogen) are treated according to manufacture's manual and mixed with approx. $1x10^{11}$ copies of the uracil-containing, 5'-biotin-terminated DNA-probe and incubated on a roller for 3 h at room temperature. Then, mixture is treated according to manufacture's manual, mixed with 500 μl DNA solution (1 M NaCl, Chapter 5.2.12.3) and incubated on a roler over night at 60°C. After incubation, mixture is resuspended in 1x NEB 4 buffer (New England Biolabs).

5.2.12.5 *Taq*I digestion

10 μg BSA (NEB) and 100 U *Taq*I (NEB) is mixed with hybridized DNA and incubated at 65 °C for 1 h, 95 °C for 5 min and 50 °C for 3 min. This digestion is repeated four times to improve digestion efficiency. After *Taq*I digestion 10 U UDG (uracil DNA glycosylase, New England Biolabs) are added and the volume adjusted to 200 μl with 1x NEB 4 buffer and incubated for 2 h at 37 °C to digest uracil-containing probe avoiding bias in following PCRs.

5.2.12.6 Copy number quantification – competitive PCR

The copy number is determined by competitive PCR (primer forward: 5'-CTC TGG GAG GAG GGG TTA AG-3', primer reverse: 5'-TCC CAA CCT CGT GAT CCG CCT-3') using internal DNA standard (Tables 24 and 25), which is amplified with the same primers as the target sequence (250 bp) but is shorter in length (180 bp). Competitor DNA (internal standard) is generated as follows: DNA is amplified using the respective forward primer and the following linker primer: 5'-TCC CAA CCT CGT GAT CCG CCT AGC CTC TGT AAG CT-3'. The PCR product is purified by agarose gel electrophoresis, isolated using Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and used as template in the subsequent amplification reaction using normal primer set. The Competitor DNA is isolated and purified as described above and quantified fluorimetrically using SYBRgreen. The competitive PCR is performed with an increasing amount of internal standard compared with constant amount of the copy number product (Table 24). PCR products are pre-stained with SYBR-Green and separated on an 3% agarose gel (Chapter 5.2.13). DNA is visualized by UV excitation and the fluorescence intensities of the PCR product of the competitor (180 bp) and of the target (250 bp) are quantified digitally (ChemiDocTM XRS+, BIO-RAD Laboratories GmbH) and analyzed with Molecular Imager® (BIO-RAD Laboratories GmbH, Chapter 5.2.13.1). The amount of target DNA is then determined by linear regression of the calibration curve.

Table 24: Pipetting scheme for copy number detection using competitive PCR. Final volume 25 μl. Internal standard (competitive DNA) concentrations (copies/μl PCR): 3.2×10^6, 1.6×10^6, 8×10^5, 4×10^5, 2×10^5, 1×10^5, 5×10^4. Dilutions of competitive DNA are performed in ncf-water.

Component	Final concentration	Volume (μl)
10x PCR Buffer (Sigma-Aldrich)	1x	2.5
$MgCl_2$	2.5 mM	2.5
Primer mix	400 nM	2
dNTP mix	0.2 mM	0.5
Template	-	1
Internal standard	-	5
Taq polymerase (Sigma-Aldrich)	0.025 U/μl	0.125
Ncf-water	-	11.375

Table 25: Temperature program of copy number quantification using competitive PCR.

Repetition	Step	Temperature (°C)	Time
1x	Denaturation	95	2 min
32x	Denaturation	95	30 sec
	Annealing	58.1	30 sec
	Extension	72	1 min
1x	Extension	72	5 min
	Cooling	10	∞

Fluorimetric quantification of DNA standards

The DNA standards for copy number quantification are quantified using fluorescence spectrometry. DNA concentrations of the DNA-standard are determined using SYBR-Green and a calf thymus (ct)-DNA standard curve as reference. ct-DNA concentration is determined photometrically (Chapter 5.2.8) and is diluted to a stock solution of 24 ng DNA/ml TE-buffer. This stock solution is further diluted to final concentrations for calibration between 0.4 and 2.4 ng DNA per ml TE-buffer (Table 26).

Table 26: Pipetting scheme for dilutions of ct-DNA Stock solution (24 ng/ml) for ct-DNA calibration. conc, concentration.

Dilution	ct-DNA stock solution (μl)	TE-buffer (μl)	Final conc ct-DNA (ng/ml)
1	10	590	0.4
2	20	580	0.8
3	30	570	1.2
4	45	630	1.6
5	60	660	2.0
6	100	900	2.4

The PCR purified DNA standard is diluted in TE-buffer to a DNA concentration of about 1.2 ng/ml. For the fluorimetrical determination 100 μl of DNA standard or ct-DNA solution 1-6 are added to 100 μl SYBR-Green V3 in triplicate in a 96-well plate. In order to ensure the accuracy of the determination, internal standards are made. Therefore 175 μl DNA standard are added to 175 μl ct-DNA dilution 4 (IS-solution). 100 μl IS-Solution are added to 100 μl SYBR-Green V3 and determined by fluorescence spectrometry (Table 27).

Table 27: I-control software settings for the fluorimetric determination of DNA in the Tecan F200 plate reader (Tecan).

Setting	Time/value
Shaking	1 min/ 1mm linear
Waiting time before measurement	4 min
Absorption wavelength	485 nm
Emission wavelength	535 nm
Gain	64
Number of flashes	20
Integration time	20 μs

5.2.12.7 Copy number quantification – TaqMan® probe-based real time PCR

The copy number quantification using TaqMan® probe-based real time PCR is performed using primer pair amplifying a short product (134 bp) and a matching TaqMan® probe (Table 28) to amplify the target sequence (Tables 29 and 30). For external calibration, the DNA probe, used for enrichment of the target sequence (Chapter 5.2.12.1), is used as DNA standard template for the calibration curve. The DNA probe is synthesized using the same PCR conditions, but without uracil (Chapter 5.2.12.1). Then, the PCR product is quantified by fluorescence spectrometry (Chapter 5.2.12.7), aliquoted and stored at -20°C.

Table 28: Primers and TaqMan® probe for real time PCR detection of TP53 copy number. BHQ1, Back Hole Quencher 1®

Forward primer	5'-GCCCTCCAGGTGAGCAGTAG-3'
Reverse primer	5'-AGGCCCTTAGCCTCTGTAAGCT-3'
TaqMan®probe	5'-FAM-ACCTCCCTATAACCCCATGAGATGTGCAA-BHQ1-3'
Length product	134 bp

Table 29: Pipetting scheme for copy number detection using TaqMan® probe-based real time PCR. Final volume 20 μl. Template, standard or sample.

Component	Final concentration	Volume (μl)
TaqMan® Gene expression master mix	1x	10
Template	0.48 pg/μl	5
Primer mix	500 nM	2
TaqMan® probe	100 nM	1
Ncf-water	-	2

Table 30: Temperature program of copy number quantification using TaqMan® probe-based real time PCR.

Repetition	Step	Temperature (°C)	Time
1x	Denaturation	95	2 min
40x	Denaturation	95	30 sec
	Annealing	60	30 sec
	Extension	72	1 min
1x	Extension	72	5 min
	Cooling	10	∞

5.2.12.8 Mutation detection

After cleavage of the target DNA sequence with *Taq*I the number of mutants are detected by means of PCR using a primer set (forward: 5'-CTC TGG GAG GAG GGG TTA AG-3', reverse: 5'-GAT CTC AGC TCA CTG CAA GCT-3') spanning the *Taq*I restriction site. The *Taq*I-digested samples are split to 79 PCRs, each reaction containing approx. $7x10^4$ total copies (mammary gland) or $7x10^3$ (cultured cells) resulting in only 8-9 positive PCRs, meaning, mutant copies (Table 32). PCR products are pre-stained with SYBR-green and separated on an 3% agarose gel (Table 34). DNA is visualized by UV excitation. If the number of mutants is more than 10 in 79 PCRs, a lower total copy number per reaction is amplified, to ensure only one mutant copy per reaction, due to Poisson distribution. Then, mutants (positive PCRs) are again digested with *Taq*I resulting in cleavage of wild-typ PCR products to avoid false positive results. False positive mutants are excluded from SMF calculation.

Table 31: Pipetting scheme for mutant detection PCR. Final volume 10 μl.

Component	Final concentration	Volume (μl)
Reaction buffer	1x	5
dNTPs	400 μM	0.625
DNA Polymerase	1.25 U	0.25
Primer mix	300 nM	1.5
Template	2 ng/μl	1
Ncf-water	-	15.125

Table 32: Temperature program of mutant detection PCR.

Repetition	Step	Temperature (°C)	Time
1x	Denaturation	95	15 min
47x	Denaturation	95	30 sec
	Annealing	60	30 sec
	Extension	72	40 sec
1x	Extension	72	5 min
	Cooling	10	∞

5.2.12.9 Calculation of mutation frequency

Mutation frequency of the target sequence is calculated with following formula:

$F = \frac{A}{(K*m*n)}$

F: mutation frequency per bp;

A: number of positive wells;

K: copy number of target DNA per mutation detection;

m: number of bases in restriction site of *Taq*I m = 4;

n: number of mutation detections

5.2.13 Agarose gel electrophoresis

Agarose gel electrophoresis is used to separate PCR products according to their product length. The components for gel electrophoresis are mixed in a 96-well plate (Table 33). From this mixture 10 μl are loaded on the agarose gel.

Table 33: Pipetting scheme for agarose gel electrophoresis.

Component	Volumes DNA ladder (μl)	Volumes sample (μl)
SYBR Green	4	2
Loading dye	2	2
DNA-ladder	1	-
PCR product	-	10
Ncf-water	10	-

For gel electrophoretic separation, the gel is run for 1 h (Conditions for each applications, Table 34). After gel electrophoresis, gels are documented using the ChemiDoc™ XRS Gel Documentation system (Biorad). The gel is exposed under ultraviolet light for 2-4 sec. The exposure time is optimal, if no overexposed details (coloured in red) can be seen in the picture.

Table 34: Conditions of agarose gel electrophoresis for specific applications. FS, field strength. NEB, New England Biolabs GmbH.

Application	Agarose (%)	DNA Ladder	FS (V/cm)
RMCA: DNA-probe (Chapter 5.2.12.1)	1.5	100 bp DNA Ladder (NEB)	5
RMCA: Copy number (Chapter 5.2.12.7)	3	50 bp DNA Ladder (NEB)	5
RMCA: Mutant detection (Chapter 5.2.12.8)	3	100 bp DNA Ladder (NEB)	5
RFL PCR (Chapter 5.2.11.2)	3.5	50 bp DNA Ladder (NEB)	5
CNP GSTT1 (Chapter 5.2.11.3)	1.5	GeneRuler™ 1kB (Thermo Scientific)	7

5.2.13.1 Densitometrical evaluation of agarose gels

In order to quantify fluorescence intensities of the bands densitometrical evaluation of agarose gels are performed. The fluorescence intensities of bands are determined using Quantity ONE software. First, all lanes on the gel are defined as "lanes" in the menu "lane tools". After that, bands are automatically identified and fluorescence intensity is quantified using "band tools". Quantification results of the bands are exported as Excel-file.

5.2.13.2 PCR product purification

Purification of PCR products are performed either from PCR-product-solution or from agarose gel slices using Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) according to manufacture's protocol.

5.2.14 E2 and E1 breast tissue level

E2 and E1 levels are determined in human mammary gland of 30 samples by means of gaschromatography (GC)/masspectometry (MS)/MS in the working group Lehmann.

5.2.15 Statistics

All statistical tests are carried out using OriginPro 8.6G or R statistical software (Chapter 5.1.7). All data are graphically checked on normal distribution and variance homogeneity. For comparison of two samples the decision tree is used (Figure 10).

Table 35: E2 and E1 levels in human mammary gland homogenates (fmol/g tissue). Determined in working group by means of GC/MS/MS. The sum of E2+E1 and the ratio of E2/E1 are calculated.

Sample #	E2	E1	Sum (E2+E1)	Ratio (E2/E1)
	(fmol/g human mammary gland tissue)			
0	0.72	0.85	1.57	0.85
1	0.19	0.38	0.56	0.50
2	0.35	1.23	1.58	0.28
3	0.55	0.86	1.41	0.64
4	0.78	13.35	14.13	0.06
5	0.53	0.60	1.13	0.88
6	0.56	0.58	1.14	0.97
7	0.07	0.54	0.61	0.13
8	1.34	1.73	3.07	0.78
10	1.24	0.34	1.59	3.62
11	0.46	1.04	1.50	0.45
12	0.80	2.53	3.33	0.32
13	1.01	1.17	2.18	0.86
14	0.07	0.32	0.39	0.21
15	0.05	0.29	0.34	0.16
16	1.47	2.72	4.19	0.54
17	0.10	0.30	0.40	0.34
18	3.47	4.97	8.44	0.70
19	0.05	0.10	0.14	0.52
20	0.10	0.64	0.74	0.16
21	0.51	0.56	1.07	0.90
24	0.12	1.38	1.50	0.08
25	1.14	1.72	2.86	0.66
26	0.94	0.85	1.78	1.11
27	0.15	0.40	0.55	0.37
28	0.81	1.49	2.29	0.54
29	1.04	0.97	2.00	1.07
30	1.12	0.60	1.72	1.86
31	0.07	0.55	0.62	0.14
32	0.63	1.24	1.87	0.51

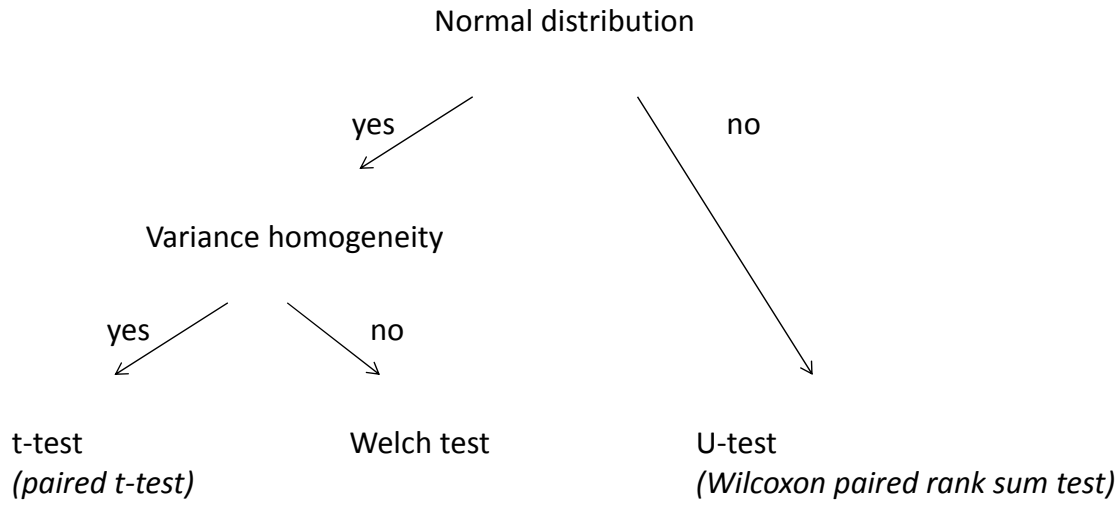


Figure 10: Decision tree for two-sample comparison. All data is graphically checked on normal distribution and variance homogeneity.

For comparison of more than two samples ANOVA (normal distributed data) or Kruskal-Wallis test (not normal distributed data) are used. For multi comparison of two samples p values are adjusted using the method of "Holm" (R statistical software, Chapter 5.1.7). For more comprehensive statistical analysis such as correlation analysis (Chapter 5.2.16) or principle component analysis (Chapter 5.2.17) R statistical software is used.

5.2.16 Correlation and cluster analysis

Correlation and cluster analysis are performed using R (Chapter 5.1.7) and the following R-script for correlation analysis, compiled with assistance of Prof. Ickstadt (Chair of Mathematical Statistics with Applications in Biometrics, University of Dortmund) and co-workers especially Helene König (Dipl. Statistician) and Claudia Köllman (Dipl. Statistician).

For complete R-scripts of all correlation and cluster analyses see enclosed CD-ROM.

Correlation analysis:

```
KorGen <- cor(dat_alles[,c(15,17,18,20:26,29:33, 35:38,40,41,43:71)] ,
+ use="pairwise.complete.obs", method = "spearman")
pdf("heatmap_Genprodukte.pdf", width=10, height=10)
heatmap.2(KorGen*(-1),
          Rowv=NA, Colv=NA, dendrogram="none",
          col=mypalette,
          colsep=1:25,rowsep=1:25,sepwidth=c(0.02,0.02),
##          cellnote=round(KorGen,2),notecex=1, notecol="white",
          breaks=c(1,0.8,0.6,0.4,0.2,0,-0.2,-0.4,-0.6,-0.8,-1)*(-1),
          trace="none", key=TRUE, keysize=1,margin=c(7,7))
dev.off()
```

5.2.17 Principal component analysis and linear regression

Principal component analysis (PCA) and linear regression are performed using R statistical software (R-script of PCA see below). For complete R-scripts of linear regression models see enclosed CD-ROM. PCA and linear regression are compiled with assistance of Helene König (Dipl. Statistician) and Claudia Köllmann (Dipl. Statistician).

```
gene2 <- data.frame(dat_alles[,"CYP1A1"],dat_alles[,"CYP1B1"],dat_alles[,"UGT1A1"],
dat_alles[,"UGT1A8"],dat_alles[,"UGT2B7"],dat_alles[,"SULT1A1"],dat_alles[,"SULT1A2"],
dat_alles[,"SULT1E1"],dat_alles[,"STS"],dat_alles[,"CYP19A1"],dat_alles[,"HSD17B1"],
dat_alles[,"HSD17B2"]

colnames(gene2) <- c("CYP1A1","CYP1B1","UGT1A1","UGT1A8","UGT2B7","SULT1A1","SULT1A2",
"SULT1E1","STS","CYP19A1","HSD17B1","HSD17B2")

index2 <- apply(gene2,1,function(x){!any(is.na(x))})

gene23_2 <- gene2[index2,]

pca <- prcomp(gene23_2, scale.=TRUE, retx = TRUE)

pdf("Screeplot_E2Gene.pdf", width=10, height=10)
screeplot(pca,type="lines")
dev.off()

plot(prcomp(gene23_2))
summary(prcomp(gene23_2, scale = TRUE))
biplot(prcomp(gene23_2, scale = TRUE))

round(pca$rotation[,1:3],2)

dat_alles$G21[index2] <- pca$x[,1]
dat_alles$G22[index2] <- pca$x[,2]
dat_alles$G23[index2] <- pca$x[,3]
```

The appropriate number of principal components are identified using scree plot. This plot graphs the eigenvalues against the component number. The "elbow point" is the point at which the remaining eigenvalues are relatively small and about the same size.

The linear regression backward selection of the linear regression models are based on Aikake information criterion (AIC).

5.2.18 Calculation of Hardy-Weinberg Equilibrium

Hardy-Weinberg Equilibrium (HWE) is calculated as follows, using the example of COMT Val108/158Met Polymorphism. Hardy-Weinberg equation:

$$p^2 + 2pq + q^2 = 1$$

Frequency of Val/Val genotype = p^2

Frequency of Val/Met genotype = 2pq

Frequency of Met/Met genotype = q^2

Using the Hardy-Weinberg equation allele frequencies are calculated:

Frequency of Val = p^2 + 0.5 (2pq)

Frequency of Met = q = 1 - p

Using the calculated allele frequencies the expected genotype frequencies are calculated:

Val/Val = p^2

Val/Met = 2pq

Met/Met = q^2

Then, the individuals, observed and expected, of each genotype are statistically compared using chi-squared test:
Calculation of χ^2:

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

The critical value for the chi-square in this case (two degrees of freedom, significance level = 0.05) is 5.991. If the calculated values are greater than the critical value, the null hypothesis is rejected and the determined frequencies are not in HWE.

5.2.19 Metabolic Network

The metabolic network is constructed with support of Prof. Dandekar (Department of Bioinformatics, University of Würzburg) and co-workers (especially Dr. Alexander Cecil). All methods and resources are used as described in Cecil et al. (2011). The pathways of the energy metabolism (pathways of TCA cycle, oxidative phosphorylation and pentose phosphate pathway and glycolysis) are used as described in Cecil et al. (2011). The network is set up using YANAsquare software (Schwarz et al., 2007). As cofactors nicotinamide adenine dinucleotide phosphate (NADPH, CYP and NQO1 reactions), S-adenosylmethionine (SAM, COMT reactions), GSH (conjugation with GSH), uridine diphosphate glucuronic acid (UDPGA, UGT reactions) and 3'-Phosphoadenosine-5'-phosphosulfate (PAPS, SULT reactions) are considered.

YANAsquare provides two opportunities to constrain the amount of metabolites and cofactors in the network: internal and external. When setting a metabolite as "internal" the production of it within the network is considered when calculating connected reactions. Thus, these metabolites/cofactors can be limited and influence the whole network. In contrast, an "external" metabolite/cofactor is unlimited available for the calculation of the fluxes. A list of all internal and external metabolites/cofactors is given in Table 36.

For calculation of network fluxes YANAsquare and a custom-made routine written in R are used (summarized in Cecil et al., 2011). As flux constraints (input data) mRNA level quantified in this work and data of RNA sequencing of (Rowley et al., 2011) are used. The calculation method is convex basis-based (Chapter 2.3.1).

Table 36: Settings for cofactors and metabolites of E1 and E1 metabolism. "external", unlimited resource. "internal", internal production is considered.

cofactor/metabolite	Setting
NADPH	external
SAM	internal
GSH	internal
UDPGA	external
PAPS	internal
E2	external
E1	external
Catechols	internal
Quinones	internal
Sulfates	external
Glucuronides	external

6 Results

In order to investigate the association of factors influencing the breast cancer risk on individual genotoxic stress caused by E1/E2 metabolism (Chapter 2.2) sufficient amount of normal human tissue of as many donors as possible was needed. At first, the tissue was characterized concerning morphology and genotypes of the tissue samples (Chapter 6.1). Then, transcript level quantification was established (Chapter 6.2) and the transcript levels in homogenate (Chapter 6.3) and epithelial cells (Chapter 6.4) determined. Then, the extent of genotoxic stress within the mammary gland tissues was assessed by DNA adduct fluxes calculated by bioinformatical network analysis (Chapter 6.5.1). Furthermore, the suitability of the Random Mutation Capture assay to determine frequencies of spontaneous mutations was investigated (Chapter 6.5.2). To investigate the influence of polymorphisms and lifestyle factors on genotoxicity and to answer the question if the E2 metabolism is decisive for the E2 mammary gland tissue levels, a correlation analysis and linear regression analysis was performed (Chapter 6.6).

6.1 Characterization of mammary gland samples

Normal human mammary gland tissues were obtained from women who underwent a mamma reduction surgery for cosmetic reasons (Chapter 5.2.2). The amount of tissue taken from one breast varied between 50-400 g. Usually, adipose tissue was the predominant tissue removed during surgery. After the tissue was removed by the surgeon it was put in ice cold physiological buffer (Chapter 5.2.2) to slow down RNA degradation. Then, as quick as possible, pure mammary gland tissue (white firm tissue) was cut out of the surrounding fatty tissue (soft yellow tissue, Figure 11). It was tried to remove as many fatty tissue as possible from the mammary gland tissue, but a complete macroscopic fat free tissue aliquot of mammary gland was gained only in some cases, strongly depending on the individual breast tissue sample composition. Usually, 2-20% fat was attached to the tissue aliquotes used for analysis. Only tissue aliquotes which were taken and flash frozen in liquide nitrogen within 10-15 min were used for RNA analysis. The later aliquots (up to 25 min after removing) were used for DNA analysis.

In this work 30 human mammary gland samples were collected (Table 37). In order to get information on all factors influencing breast cancer risk by interaction with E1/E2 metabolism (Chapter 2.2), factors, which could not assessed experimentally, were assessed by a questionnaire answered by the female donors (Chapter 6.1.1). Furthermore, the morphology (Chapter 6.1.2) and the genotype of each sample (Chapter 6.1.3) were determined experimentally.

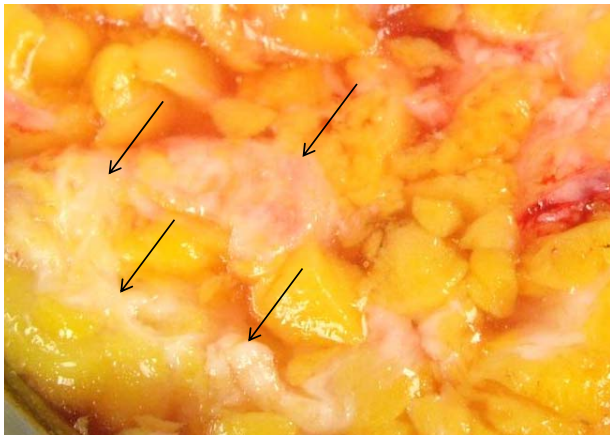


Figure 11: Photo of "white" human mammary gland tissue (black arrows) and surrounding "yellow" adipose tissue in ice-cold physiological buffer.

6.1.1 Questionnaire

First, factors were identified influencing breast cancer risk in Caucasian women by possibly interfering with E1/E2 biosynthesis and metabolism (BMI, age, parity, smoking habits, alcohol consumption, Chapter 2.4). Then, all information concerning the factors and the intake of hormone active drug or dietary supplements, interfering with E2 tissue or blood levels and E2 response, were compiled (Chapter A9.1). To get information concerning BMI the female donors were asked about their hight and weight. Lobule types were assessed by including age and parity in the questionnaire and histological evaluation of mammary gland morphology (Chapter 6.1.2).

Additionally, information concerning factors such as smoking habits, alcohol consumption and intake of hormone active drugs or dietary supplements (Chapter 2.2) were answered by the female donors.

6.1.2 Morphology

The morphology of the mammary gland is depending on age and parity (Chapter 2.2). In breast tissue of nulliparous women the lobule type 1 and terminal ducts, occasionally lobule type 2 or 3 are observed constant through lifespan (Chapter 2.1). In mammary gland of parous women the lobule types 2 and 3 are predominant during reproductive years, but decreasing back to lobule type 1 after fourth decade of life (Chapter 2.1). The age-dependent involution of mammary gland tissue of parous women has been described to pre-date the onset of menopause (Chapter 2.1).

In order to investigate if the lobule type of parous women investigated in this study was already decreased back to lobule type 1 (lobule type 1_{parous}), the tissue of the respective women were investigated histochemically.

Table 37: Distribution of the 30 mammary gland samples collected in the present study concerning age, BMI, number of pregnancies (PREG), smoking habit (Smoke), intake of hormone active drugs (HAD) or dietary supplements (HDS) and alcohol consumption (ALC) of their donors. n.i., no information.

Factor	Categories	# of women
Age	<40	12
	40-50	7
	>50	11
BMI	<25	14
	>25	16
PREG	0	12
	1	4
	2	8
	3	3
	n.i.	3
Smoke	yes	6
	no	23
	n.i.	1
HAD	yes	7
	no	23
HDS	yes	1
	no	29
ALC	no	12
	20-50 g	10
	50-100 g	3
	>100 g	2
	n.i.	3

First, the samples were pre-selected by age, every parous women over 40 years (10 samples) were analyzed regarding their lobule type (Chapter 6.4.1). Mammary gland tissue of lobule type 1 is characterized by mostly ducts and only few ductules per lobule (11±6) and tissue of lobule type 3 is dominated by lobules containing 81±17 ductules (Russo and Russo, 2004, Figure 12). Eight of the ten mammary gland tissue samples investigated in this work were lobule type 1_{parous} and two were lobule type 2/3 (Chapter A9.2).

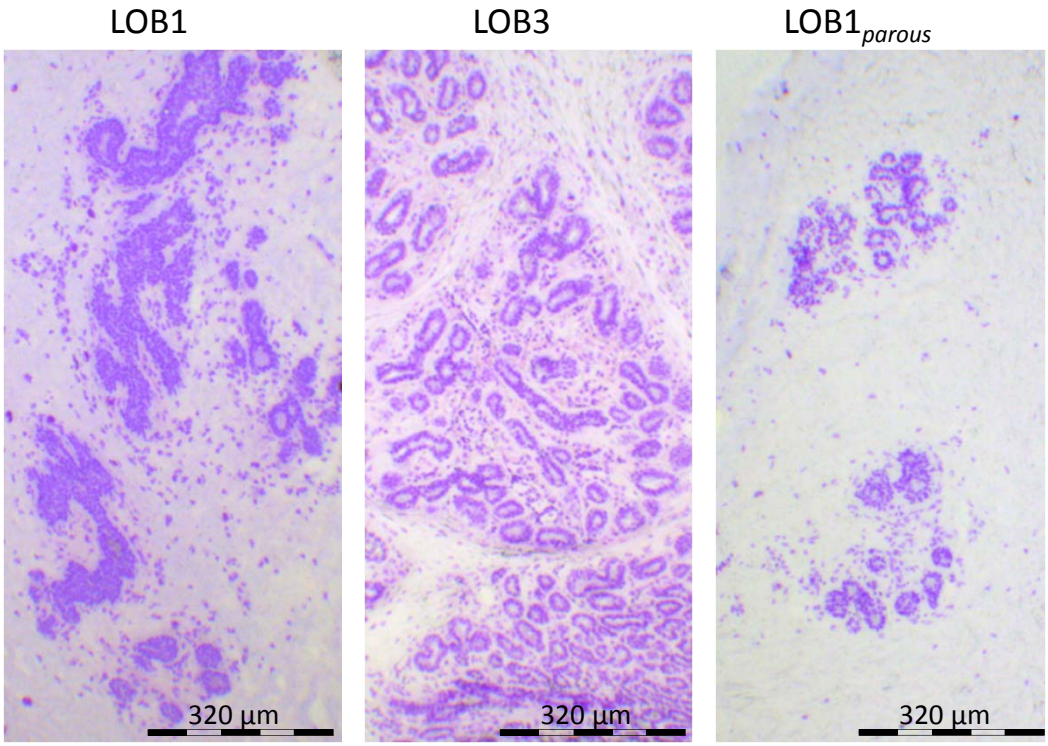


Figure 12: Representative histological micrograph of a young nulliparous woman with lobule type (LOB) 1 (sample #1), of a parous woman at age of 32 with LOB3 (sample #29) and a parous woman at age of 51 with LOB1$_{parous}$ (sample #5). Sections were stained with cresyl violet (Chapter 5.2.4.1). Magnification 6.3x.

6.1.3 Polymorphisms

In order to determine the genotype of the mammary gland tissue samples collected in the study, it was focused on genetic polymorphisms in genes encoding enzymes involved in E2 metabolism associated with increased breast cancer risk (GSTT1, NQO1, HSD17B1, CYP1B1, GSTM1, Chapter 2.2). Moreover, only polymorphisms, where the genotype correlated with breast cancer has a frequency of at least 10% in caucasian population, were taken into account (Chapter 2.4.1). In addition to the clearly breast cancer risk associated polymorphisms in GSTT1 CNP and NQO1 Pro187Ser the COMT Val108/158Met polymorphism was investigated, since an epidemiological study investigating the interaction of COMT Val108/158Met and NQO1 Pro187Ser has been lacking up to now. The polymorphism in HSD17B1 was not further investigated in this work, because is it not directly involved in metabolism of reactive E1/E2 metabolites. The investigation of CYP1B1 Leu432Val was also not addressed in this study, because it was focused on the most significantly correlated polymorphisms with breast cancer risk. The investigation of the GSTM1 CNP was omitted from this work, since the expected transcript levels were extremely low

compared to the transcript levels of GSTT1. Therefore, it was focused on GSTT1 CNP, because more impact on the E1/E2 metabolism was expected. Therefore, the aim was to establish reliable methods for determination of the polymorphisms COMT Val108/158Met (rs4680), NQO1 Pro187Ser (rs1800566) and GSTT1, one method for the detection of single nucleotide polymorphisms (SNPs, Chapter 6.1.3.1) and one for copy number polymorphisms (CNPs, Chapter 6.1.3.2).

6.1.3.1 Method establishment and application for the detection of SNPs

SNPs are characterized by a nucleotide exchange in a gene, causing a change in amino acid sequence (Table 3). There are many methods to determine SNPs. Generally the discrimination reaction between the two alleles can be divided in five types of assays (i) hybridization-based, (ii) primer extension-based, (iii) allele specific oligonucleotide ligation-based, (iv) invasive cleavage-based (summarized in Sobrino et al., 2005) and a very recent approach (v) through next generation sequencing (Kumar et al., 2012). For these discrimination reactions there are various detection methods available: (i) electrophoresis, (ii) fluorescence resonance energy transfer, (iii) fluorescence polarization, (iv) arrays, (v) mass spectrometry and (vi) luminescence (summarized in Sobrino et al., 2005). The aim was a routine genotyping of two SNPs in all women in this project, so, a quick, flexible and resource-saving method was required. The hybridization-based allelic discrimination assay meets all this requirements. It is a commercial available assay for determination of most common human SNPs in a micro-well-plate format. However, for reliable determinations a positive control of each genetic variant is necessary, since theoretically, (i) the fluorescence intensity for the unmatched probed is not necessarily zero, (ii) the maximum of both dyes are not necessarily equal and (iii) absolute intensities may vary from PCR run to PCR run. Thus, the absolute position of each genetic variant may shift in the cluster plot. Since DNA of cultured cells is of unlimited availability, cell lines are good positive controls for each PCR run. Therefore, a second well established method, the restriction fragment length (RFL) PCR (Chapter 6.1.3.1.1), was used to determine the genotypes of seven commonly used human cell lines (MCF-7, HT-29, HelaS3, Ishikawa, human chorion villi cells (HCV), human skin fibroblasts (AG0) and MDA) for the SNP COMT Val108/158Met and results were compared to results of allelic discrimination assay (Chapter 6.1.3.1.2). Then, positive controls for the determination using allelic discrimination assay of the SNP NQO1 Pro187Ser were identified (Chapter 6.1.3.1.3).

For both methods, total DNA of each human cell line was isolated by chloroform extraction and precipitation with 2-propanol (Chapter 5.2.7).

6.1.3.1.1 RFL PCR COMT Val108/158Met

The RFL PCR method was used according to Sazci et al. 2004 and Yim et al. 2001(Chapter 5.2.11.2). It is based on the different fragment length of the two possible allelic variants after restriction digest of the PCR product with the restriction enzyme *Nla*III (restriction site CATG). For the Val allele, fragments with the length of 114 bp, 83 bp and 20 bp and for the Met allele (two restriction sites), fragments with the length of 96 bp, 83 bp, 20 bp, 18 bp are generated (three restriction sites, Figure 13A). The heterozygous variant generates all mentioned fragments (Figure 13A).

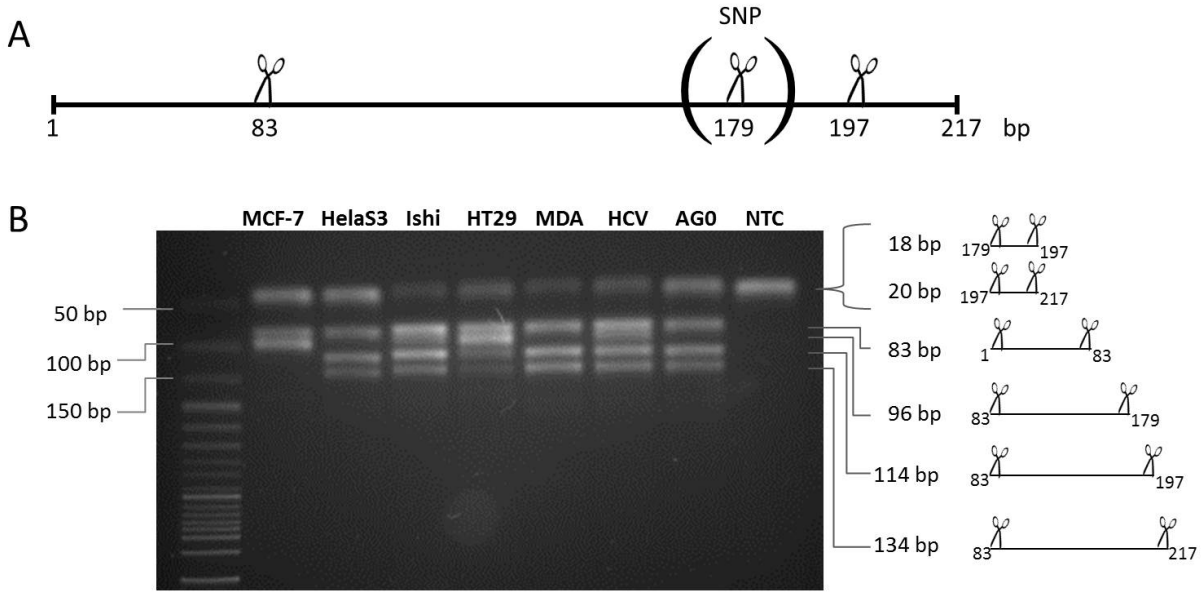


Figure 13: (A) Scheme of the amplified PCR products of RFL PCR for genotyping COMT Val108/158Met before and after restriction digest with *Nla*III. Before resticition digest all three genetic variants give the same PCR product (217 bp). After restriction digest fragments with the length of 114 bp, 83 bp, 20 bp and 96 bp, 83 bp, 20 bp were observed for the Val/Val and Met/Met variant, respectively. All fragments are detected for heterozygote (Val/Met). (B) Electrophoretic separation (5 V/cm, 1.5 h, 3.5% agarose gel stained with Sybr Green) of *Nla*III digested PCR products (RFL PCR) of the human cell lines MCF-7, MDA, HelaS3, Ishikawa, HT-29, HCV and AG0. Ishi, Ishikawa cells. NTC, no template control.

For fragment length determination the digested PCR products were separated by means of agarose gel electrophoresis (Chapter 5.2.11.2). For the MCF-7 cell line bands at approx. 20 bp, 80 and 95 bp were observed (Figure 13B), indicating the homozygous Met/Met variant, in accordance with (Dawling et al., 2001). The missing band at 18 bp is inseparable from the band at 20 bp and since a band of the same length was observed in the no template control, both bands have the same length at which primer and/or primer dimer bands are usually detected. For HelaS3 cells (genotyped as Val/Val with allelic discrimination) four

bands were detected at approx. 20 bp, 80 bp, 110 bp and 135 bp (Figure 13B), indicating the Val/Val genotype. The band at 135 bp was observed for all heterozygous and homozygous (Val/Val) cell lines (Figure 13B), but was not observed by Sazci et al. (2004) and Yim et al. (2001). In order to verify the amplification of the correct product, the sequence of the uncut PCR product (template DNA of Ishikawa cells, heterozygous) was investigated by Sanger sequencing (Chapter 5.2.11.2) and compared to the gene bank sequence (ENSG00000093010, Ensembl genome database). The product (217 bp) was double sequenced from position 44 to 160 (both DNA strands) and single sequenced from position 5 to 206, which includes all predicted NlAIII restriction sites (Figure 13A). The alignment with the Ensembl database sequence revealed one mismatch at bp position 177 in forward primer sequence. This was expected, since this is the location of the SNP and Ishikawa cells are heterozygous. So, 200 bp of the 217 bp of the amplification product were verified and revealed no hint at the additional restriction fragment of 135 bp. The only possibility is a incomplete restriction digest at bp position 197 which would result in 134 bp fragment and not in 114 bp and 20 bp fragments. The 135 bp fragment is not generated in the Met/Met genotype, since this it also cut at the SNP position generating a 96 bp and 38 bp fragment. The 38 bp fragment is then cut into 20 bp and 18 bp fragments (bp position 197 of the uncut product, Figure 13A).

Interestingly intensities of band derived from HT29 cells exhibited different intensities than Ishikawa and HCV cells in agarose gel (Figure 13B), although all three cells lines were heterozygote.

HT-29 are known to be hyper-triploid (Kleivi et al., 2004), so different genetic variants by loss or duplication of one allele are possible. Another possibility for more intense bands is a heterogeneous cell population with subpopulations having different genotypes. This heterogeneity in HT-29 cells is described for different morphologies and enzyme activities in differentiated cells of this cell line (Stokrova et al., 2006).

6.1.3.1.2 Allelic discrimination COMT Val108/158Met

The second method used to determine the COMT Val108/158Met SNP is the allelic discrimination assay. It is a real time PCR method using different fluorescence dyes bound to TaqMan® probes that target SNP sites (Afonina et al., 1997; Kutyavin et al., 1997). One fluorescent dye-labeled TaqMan® probe is a perfect match to the wild type allele G (Figure 14A) and the other fluorescent dye-labeled probe is a perfect match to allele A (Figure 14A). The allelic discrimination assay includes a fluorescent measurement of the intensity of both dyes immediately before and after the PCR run. Thereby, it classifies samples as homozygous (samples having only allele G or allele A) or heterozygous (samples having both allele

G and A), graphically displayed in a cluster plot (Figure 14B).

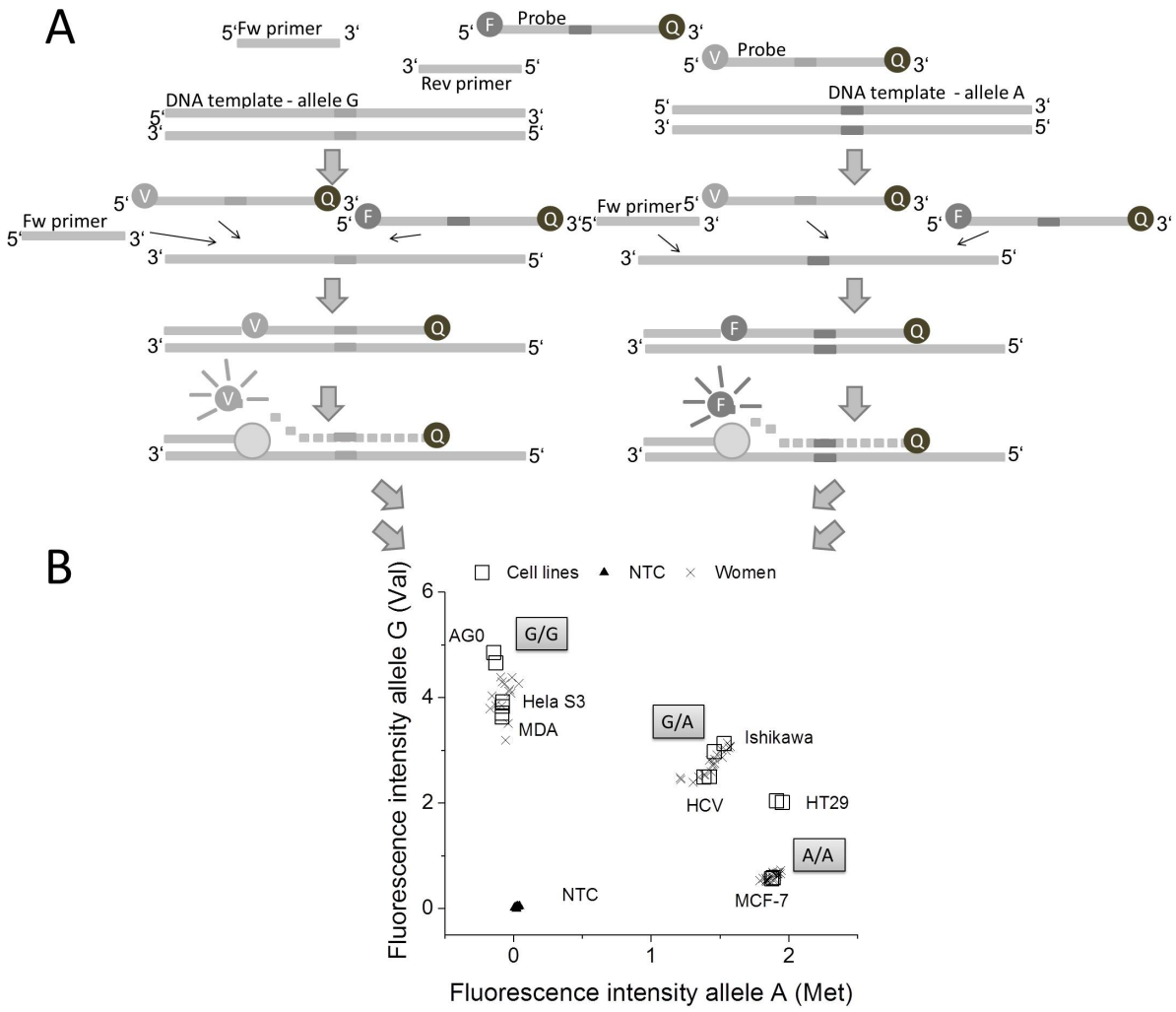


Figure 14: Scheme and results of allelic discrimination assay for COMT Val108/158Met (rs4680). (A) Simplified schematic first cycle of an allelic discrimination PCR with both fluorescence dye-labeled probes for allele G and A, primer pair and the DNA template. (B) Cluster plot of allelic discrimination for COMT Val108/158Met polymorphism of human cell lines MCF-7, MDA, HelaS3, Ishi (Ishikawa), HT-29, HCV and AG0 and all analyzed women. F, Fam™. V, VIC®. Q, quencher. Fw, forward. Rev, reverse. NTC, no template control.

For determination of COMT Val108/158Met and NQO1 Pro187Ser commercial assays are available, including specific primer and TaqMan® probes labeled with 6-carboxy-fluorescein (FAM™) and 4,7,2-trichloro-7-phenyl-6-carboxyfluorescein (VIC®) fluorescence dyes. For reliable determinations a positive control of each genetic variant is necessary, since theoretically, (i) the fluorescence intensity for the unmatched probed is not necessarily zero, (ii) the maximum of both dyes are not necessarily equal and (iii) absolute intensities may vary from

PCR run to PCR run. Thus, the absolute position of each genetic variant may shift in the cluster plot.

Since DNA of cultured cells is of unlimited availability, cell lines are good positive controls for each PCR run. Thus, the seven cell lines genotyped with RFL PCR were genotyped using the allelic discrimination assay.

The allelic discrimination assay was performed in a 384 well plate in a 5 μl reaction (Chapter 5.2.11.1). Then, fluorescence intensities of both fluorescence dyes (FAM™ allele Val and VIC® allele Met) were determined and graphically displayed in a cluster plot. For the MCF-7 cells a high fluorescence intensity of the Met allele and low fluorescence intensity of the Val allele were observed (Figure 14B), indicating MCF-7 cells homozygous for the low activity allele (Met/Met) in accordance with the result of the RFL PCR and (Dawling et al., 2001). HelaS3, AG0 and MDA clustered at a high fluorescence intensity of the Val allele and zero fluorescence intensity of the Met allele, accordingly, they were homozygous for the high activity Val/Val allele. Ishikawa and HCV were heterozygous (Val/Met) and HT-29 cells were not classifiable as heterozygous or homozygous (Met/Met) in the cluster plot (Figure 14B), comparable to results of the RFL PCR. In conclusion, the results of RFL PCR (Figure 13B) of the tested human cell lines were also obtained with the allelic discrimination assay. Hence, MCF-7 (Met/Met), Ishikawa (Val/Met) and HelaS3 (Val/Val) cells were selected as positive controls for COMT Val108/158Met genotyping by allelic discrimination. Using the positive controls in the allelic discrimination assay 30 mammary gland tissue samples were determined concerning their COMT Val108/158Met SNP (Figure 14B).

6.1.3.1.3 Allelic discrimination NQO1 Pro187Ser

For determination of the NQO1 Pro187Ser polymorphism the commercial available allelic discrimination assay was used analogous to COMT allelic discrimination assay (Chapter 5.2.11.1) with the FAM™-labeled probe for the Pro allele and the VIC®-labeled probe for the Ser allele.

HelaS3, HCV, AG0, Ishikawa and HT-29 cells (in accordance with Traver et al., 1992; Siegel et al., 2001) were homozygous for the Pro allele; MCF-7 cells were heterozygous (Pro/Ser); MDA cells were homozygous for the Ser allele (Figure 15). Thus, for the homozygous Pro allele Ishikawa, HelaS3 or HT29-cells, for the heterozygous variant MDA cells and for the homozygous Ser allele MCF-7, they were the most suitable positive controls. Using the positive controls in the allelic discrimination assay 30 mammary gland tissue samples were determined concerning their NQO1 Pro187Ser SNP (Figure 15).

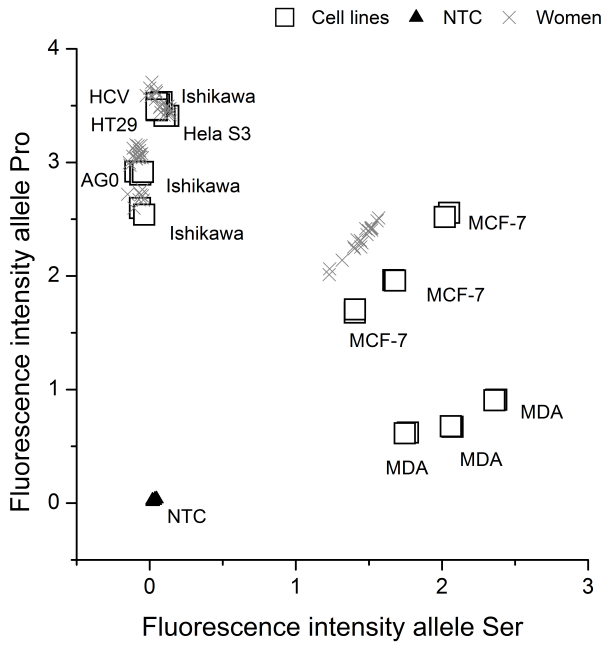


Figure 15: Cluster plot of allelic discrimination for NQO1 Pro187Ser polymorphism of human cell lines MCF-7, MDA, HelaS3, Ishikawa, HT-29, HCV and AG0 and all human mammary gland samples analyzed in the study out of three PCR runs. NTC, no template control.

6.1.3.2 Method establishment and application for the determination of CNP

The polymorphism in GSTT1 is a copy number variation, meaning the complete loss of the functional gene of both copies (nullizygous or -/-) or of one copy (hemizygous or +/-) or both copies are functional present (homozygous or +/+). The nullizygous variation results in loss of the gene (Pemble et al., 1994) and thus mRNA, whereas the hemizygous and homozygous variant cannot be clearly distinguish by the amount of mRNA. For determination of the GSTT1 CNP, a DNA based method, first described by (Buchard et al., 2007), was used. This is a multiplex PCR amplifying a PCR product with 969 bp for the present gene and a PCR product with 3106 bp for the deleterious gene (Buchard et al., 2007, Figure 16A).

The method described by (Buchard et al., 2007) was established for simultaneous genotyping of GSTT1, GSTM and GSTP (generating PCR products up to 4748 bp). Since only the GSTT1 genotyping was of interest (Chapter 2.4.1), the annealing/elongation step was reduced from 7 min to 4 min at 68 °C (Chapter 5.2.11.3). Due to changes in the method, primer pairs were at first tested in separate reaction mixtures. To generate both possible amplification products, two different breast tissue samples were used as template DNA, one with no detectable mRNA of GSTT1 (Chapter 6.3, expecting PCR product with 3106 bp) and one with detectable mRNA of GSTT1 (Chapter 6.3, expecting PCR product with 3106 bp

and/or 969 bp). For the sample with no detectable GSTT1 mRNA only a PCR product with the length of approx. 3000 bp was detected in the reaction mixtures containing both primer pairs or the primer pair amplifying the deleterious gene product (lanes B and D in Figure 16B). In the reaction mixture containing only the primer pair amplifying the present gene PCR product, no band was observed (lane C in Figure 16B). For the sample with detectable GSTT1 mRNA a band at 3106 bp and a band at 969 bp were observed in the reaction mixtures containing the primer pair amplifying the deleterious gene product and amplifying the present gene product, respectively (lanes E and F in Figure 16B). Amplification with the reaction mixture containing both primer pairs yielded two detectable PCR products with 3106 bp and 969 bp (lane G in Figure 16B). In conclusion, for both samples and for all primer pair variations the expected amplification products were detected.

To guarantee a successful multiplex PCR in every PCR run, a positive control is necessary. So, the best positive control would be a hemizygous cancer cell line. Since, there was no literature concerning the genotype of cancer cell lines available, seven cell lines were genotyped. HelaS3, Ishikawa, MDA, HCV, Ag0 were hemizygous, HT-29 cells were homozygous and MDA cells were nullizygous (Chapter 9 and Table 38). Thus, HelaS3, Ishikawa and MDA are the most suitable positive controls for this method. Using the positive controls 30 mammary gland tissue samples were determined concerning their GSTT1 CNP (Table 38).

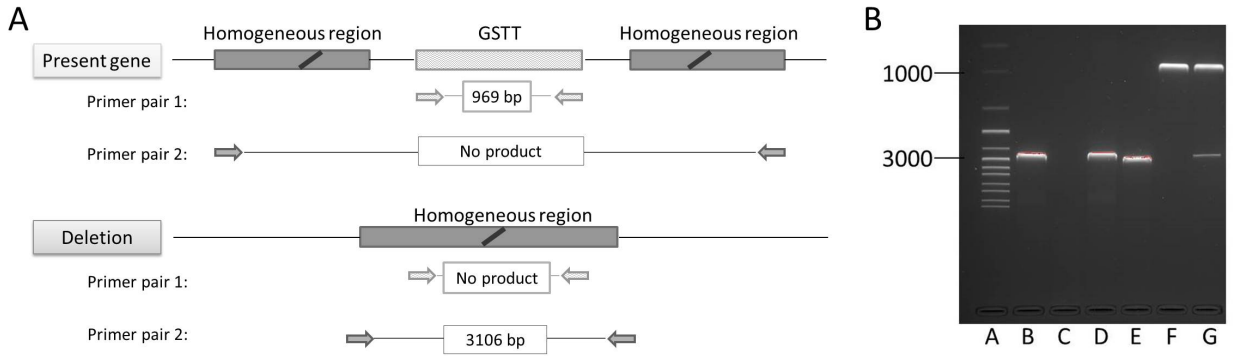


Figure 16: (A) Scheme of GSTT1 polymorphism detection. A multiplex PCR with two primer pairs in one reaction mix. Primer pair 1 amplifying the present gene PCR product (969 bp) and primer pair 2 amplifying the deleterious gene PCR product (3106 bp, modified from Buchard et al., 2007). If the GSTT1 gene is present, the product of primer pair 2 is too long to be amplified. (B) Electrophoretic separation (7 V/cm, 1 h, 1.5% agarose gel stained with Sybr Green) of PCR products of DNA of breast tissue with no detctable mRNA (lanes B-D) and DNA of breast tissue with detectable mRNA (lanes E-G). Lane A, 1 kb DNA ladder (Thermo Scientific). Lanes B and E, reactions with primer for deletorious gene product. Lanes C and F, reactions with primer for the present gene product. Lanes D and G, reactions with both primer pairs.

6.1.3.3 Genotypes of human cell lines and human mammary gland tissue samples

For the detection of all three polymorphisms (COMT Val108/158Met, Pro187Ser and GSTT1) reliable methods were established. Seven human cell lines were genotyped for the three polymorphisms. HelaS3 and AG0 had the same genotype with the highly active COMT Val/Val and NQO1 Pro/Pro variant and one copy (+/-) of GSTT1 gene (Table 38). Likewise, Ishikawa and HCV were heterozygous/hemizygous for COMT and GSTT1 and homozygous for NQO1 Pro/Pro. HT-29 cells were heterozygous for COMT polymorphisms and homozygous for the highly active NQO1 Pro/Pro and were the only tested cell line with two copies of the GSTT1 gene (+/+). The two breast cancer cell lines had different genotypes: MCF-7 were homozygous for the low activity allele COMT Met/Met whereas the MDA cells homozygous for the high activity allele COMT Val/Val. The MDA cell were the only cell line with the low activity allele NQO1 Ser/Ser and no copy of the GSTT1 gene (-/-). MCF-7 cells were the only tested cell line heterozygous for NQO1 Pro/Ser and had one copy of the GSTT1 gene (+/-, Table 38).

Up to now, of all genotyped cell lines only the previously mentioned genotypes for HT-29 (NQO1 Pro/Pro, Traver et al., 1992; Siegel et al., 2001) and MCF-7 (COMT Met/Met, Dawling et al., 2001) have been described. The GSTT1 polymorphism was determined in different hematopoeitic cell lines (Wang et al., 2000), but not in one of the tested cell lines. But, genotypes of the cell lines should be taken into account when performing investigations on E2 metabolism or other metabolisms, including one of the three enzymes, since the polymorphism results in altered enzymatic performance (Chapter 2.4.1).

Table 38: Genotypes for COMT Val108/158Met, NQO1 Pro187Ser and GSTT1 of human cell lines. Ishi, Ishikawa cells.

Cell line	COMT	NQO1	GSTT1
HelaS3	Val/Val	Pro/Pro	+/-
Ishi	Val/Met	Pro/Pro	+/-
MCF7	Met/Met	Pro/Ser	+/-
HT-29	Val/Met, Met/Met	Pro/Pro	+/+
HCV	Val/Met	Pro/Pro	+/-
MDA	Val/val	Ser/Ser	-/-
AG0	Val/Val	Pro/pro	+/-

With the three described methods 30 women were genotyped (Table 39). The Hardy-Weinberg Equilibrium (HWE) states that if two alleles, G and A, with frequencies p and q=1-p, are in equilibrium in a population, then the proportion of people with genotypes GG, GA and AA will be p^2,2pq and q^2 (summarized in Minelli et al., 2008). Departures from HWE can be due to non-random mating, selection or migration (Minelli et al., 2008). HWE

was calculated (Chapter 5.2.18) for the detected frequencies of the three polymorphisms (Appendix: Tables 53, 54, 55) and they were all in HWE (Appendix: Table 56), indicating a random selection of subjects for this study.

Additionally, the frequencies of each genotype were compared to frequencies observed in other studies among Germans (summarized in Table A57). In total, in these studies 3156, 1519, 1242 German individual were genotyped for GSTT1 CNP, NQO1 Pro187Ser and COMT Val108/158Met, respectively, indicating reliable frequencies of the German population. GSTT1 CNP homozygotes (33.3%, Table 39) were comparable to other studies, ranging from 32.6-33.5%. The frequency of the hemizygous (60.0%, Table 39) and nullizygous (6.6%, Table 39) individuals differed from the other studies by about 10% (hemizugote: 49.4-50.0% and nullizygote: 17.1-19.3%). Likewise, there were differences in the frequnecy of the low activity genotype COMT Met/Met (46.6%, Table 39) and the heterozygous Val/Met (30.0%, Table 39) compared to the listed studies, ranging from 24.0-28.0% (Met/Met) and 49.0-54.9% (Val/Met). The observed frequency for the high activity genotype Val/Val (23.3%, Table 39) was in the range (22.0-28.0%) of the other studies. The observed genotype frequencies of the NQO1 Pro187Ser polymorphism (Pro/Pro 73.3% and Pro/Ser 26.6%, Table 39) were in range of other studies (Pro/Pro 67.6-73.4% and Pro/Ser 25.0% to Pro/Ser+Ser/Ser 32.4%). The frequency of the low activity genotype NQO1 Ser/Ser was not observed in this study which is in accordance with the published data (Ser/Ser 1.6%), since this would mean "0.5" individuals of the 31 tested in this study. The discrepancies in the frequencies of the COMT SNP and the GSTT1 CNP were perhaps due to the comparativly low number of genotyped individuals.

Table 39: Polymorphism frequencies (absolute and percent) of COMT Val108/158Met, NQO1 Pro187Ser and GSTT1 CNP of 30 women. Last line, frequencies observed among Germans in other studies.

COMT			**NQO1**			**GSTT1**		
Val/Val	Val/Met	Met/Met	Pro/Pro	Pro/Ser	Ser/Ser	+/+	+/-	-/-
7 (23.3%)	9 (30.0%)	14 (46.6%)	22 (73.3%)	8 (26.6%)	0 (0%)	10 (33.3%)	18 (60.0%)	2 (6.6%)
22.0-28.0%	49.0-54.9%	24.0-28.0%	67.6-73.4%	25.0%-32.4%	1.6%	32.6-33.5%	49.4-50.0%	17.1-19.3%
Rudolph et al. (2011)Schmahl et al. (2012) Reuter et al. (2006) Majic et al. (2011)			Zhang et al. (2003) Timofeeva et al. (2010)			Rudolph et al. (2012) Bruhn et al. (1998) Timofeeva et al. (2010)		

In order to investigate whether the polymorphism has an impact on mRNA level of the transcripts, transcript levels of each genetic variant of the three genes was compared (Figure 17). As expected (Dawling et al., 2001; Siegel et al., 2001, Chapter 2.4.1), neither for the COMT SNP or the NQO1 SNP, was a significant impact of genotypes on transcript levels was observed (ANOVA, p=0.2800 and U-test, p=0.7451, respectively).

For GSTT1 CNP only +/+ and +/- were tested, since there was no detectable mRNA for the nullizygous genotype. A higher transcript level was observed for the +/+ genotype compared to the +/- genotype, although borderline statistically significant (Welch test, p=0.0531). Up to now, the impact of GSTT1 CNP on the mRNA level of GSTT1, although borderline significant, has not been shown. Only a positive correlation of the genotype with the enzyme activity was described (Sprenger et al., 2000).

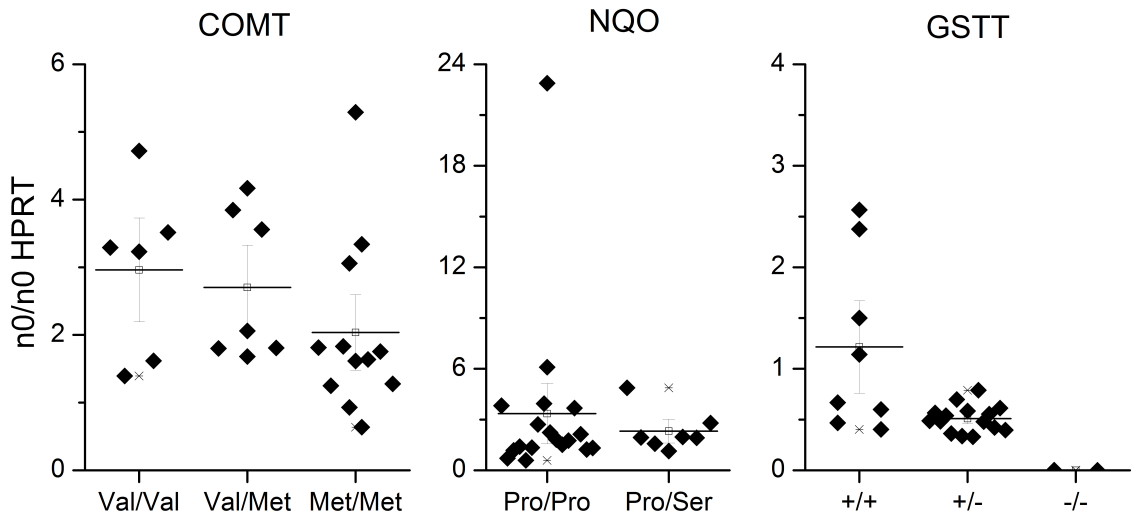


Figure 17: Boxplot of mRNA level of COMT, NQO1 and GSTT1 vs. their genotypes of 25 women. Neither for the COMT SNP or the NQO1 SNP, was a significant impact on mRNA was observed (ANOVA, p= 0.2800 and U-test, p=0.7451, respectively). For GSTT1 CNP only +/+ and +/- were tested, since there was no detectable mRNA for the nullizygous genotype. A higher mRNA level was observed for the +/+ genotype compared to the +/- genotype, although borderline significant (Welch test, p=0.0531).

6.2 Establishment of mRNA level quantification using TaqMan® Low Density Array

For the quantification of the transcript levels of genes involved in E2/E1 metabolism, apoptosis, cell cycle and paracrine signal transduction and marker genes for proliferation and ESR1 activation (Chapter 2.2) a method was chosen which is capable of quantifying as many transcripts per samples as possible within one PCR run, since more transcripts than samples were expected for the study. TaqMan® probe-based low density array technique (TLDA) met this challenge. The TLDA contains eight sample-loading slots each connected by a micro channel to 48 miniature reaction chambers (1 μl volume) for a total of 384 wells per card. Transcript specific primers and TaqMan® probes of a customized set of transcripts were factory-designed and embedded in each well (Chapter 5.1.6). 100 μl reaction mixture are loaded in each slot consisting of 50 μl master mix and 50 μl cDNA diluted with ncf-water (Chapter 5.2.10). Depending on TLDA layout chosen one to eight samples and 16-384 transcripts are quantified in triplet or single reaction (Chapter 5.1.6).

Therefore, the first step was to set up the customized TLDA with all important transcripts involved in E2/E1 metabolism, apoptosis, cell cycle and paracrine signal transduction and marker genes for proliferation and ESR1 activation (Chapter 6.2.1).

In the second step the aim was to establish the TLDA technique for a reliable and reproducible quantification of all transcript levels (Chapter 6.2.2). This includes, in the third step, the establishment of a method for quantifying very low expressed transcripts and for quantifying transcript levels in laser dissected cells (Chapter 6.4) of the mammary gland (Chapter 6.2.4). In the fourth step the influence of the sample taking site of human mammary gland was investigated (Chapter 6.2.5).

6.2.1 Selection of transcripts

The selection of the transcripts encoding enzymes involved in E1/E2 metabolism was based on two criteria: First, the isoenzymes/enzymes are metabolizing E1, E2 or their metabolites. Second, the enzymes were previously detected in normal mammary gland tissue or the data was not sufficient (Chapter 2.2).

Reference genes

Since most of the target transcript levels were expected to be very low, hypoxanthine-guanine phosphoribosyltransferase 1 (HPRT1), TATA box binding protein (TBP) and glucuronidase beta (GUSB) were chosen as reference genes, because all three correlated well with other levels of reference genes (de Kok et al., 2005) and were low expressed in mammary gland compared to other frequently used reference genes such as 18S or glyceraldehyde-3-phosphate dehydrogenase (de Kok et al., 2005).

E2 formation and metabolism

The main focus of this work was on oxidative E1/E2 metabolism, thus only the most important enzymes involved in E1/E2 formation in mammary gland (Chapter 2.2.1) were taken into account: CYP19A1 and HSD17B1/HSD17B2.

Six transcripts of SULT-family, which are known to sulfonate E1/E2 or its respective metabolites (SULT1A1, SULT1A2, SULT1A3, SULT1A4, SULT1E1 and SULT2A1, Chapter 2.2.2) were selected. For the detection of SULT1A3 and 1A4, only one assay was available (SULT1A3/4, Hs00413970_m1, Life Technology), which did not differentiate between both isoenzymes.

Since all estrogen sulfates are de-sulfonated by STS (Chapter 2.2.2), the transcript was also selected for quantification in this work.

Five of the transcripts of the UGT family, which are known to catalyze the glucuronidation of E1, E1 and the respective metabolites (UGT1A1, UGT1A3, UGT1A8, UGT1A10 and UGT2B7, Chapter 2.2.2) were selected. The transcript of UGT1A9 was not considered, since it had not been detected in normal human mammary gland tissue (Chapter 2.2.2). For the detection of UGT1A4 only one assay was available (UGT1A3/4, Hs01592480_m1, Life Technology) detecting both transcripts UGT1A4 and UGT1A4.

Since the other CYPs isoforms were either not expressed in normal human mammary gland tissue or catalyzed only the hydroxylation at position 16 (Chapter 2.2.2), only the two main CYPs isoenzymes CYP1A1 and 1B1 in normal mammary gland tissue catalyzing the hydroxylation at position 2 and 4 (Chapter 2.2.2) were considered.

The two most abundant isozymes of GSTs (GSTT1 and GSTP1, Chapter 2.2.2) were chosen, catalyzing the important conjugation of quinones with GSH (Chapter 2.2.2). Furthermore, transcripts of the two other important enzymes involved in inactivation of E1/E2-catechols and -quinones, COMT and NQO1 (Chapter 2.2.2), were investigated using TLDA.

Marker genes for cellular stress

The transcripts of GCLC and GADD45A, as well as both GSTs and NQO1 were used as stress makers (Chapter 2.3).

Transcription factors involved in E2 metabolism

Nuclear receptors known to be involved in induction of the E1/E2 metabolizing enzymes, NR1I2, NR1I3, AHR and ARNT (Chapter 2.2.3) were investigated. For characterization of estrogen receptor status, the transcript levels of ESR1 and ESR2 were determined using TLDA. The transcript levels of TFF1, CCDN1, PGR and AREG, as marker-transcripts for ligand-activated ESR1 (Chapter 2.1), were investigated. The transcript levels AREG, WNT4, WNT5A and TGFB1 were investigated as important proteins of the paracrine signal transduction (Chapter 2.1).

Proliferation marker, cell cycle and involution

For further characterization of proliferation and apoptosis, transcripts of enzymes involved in cell cycle (CDKN1A, CDKN1B, TP53, MDM2, SFN, Chapter 2.1) as well as MKI67, a marker for proliferation (Chapter 2.1), were determined using TLDA. The transcript levels of BMF, BAD and BAX were determined as important mediators of involution of the mammary gland (Chapter 2.1).

Thus, 49 transcripts met the criteria described above and were included in the TLDA. Only two TLDA formats were worth considering regarding number of transcripts: 32 transcripts in triplet (4 samples per card) or 48 transcripts in single reaction (8 samples per card). The remaining transcripts had to be quantified in additional real time PCRs in 96-well format.

6.2.2 Selection of TLDA format

In order to investigate if transcripts needed to be quantified in triplicate due to high variations, the reproducibility of triplicate per slot was determinded (Chapter 6.2.2.2). Since ingredients of cDNA synthesis are able to inhibit polymerase reaction, and a reproducible PCR reaction, first, the TLDA was characterized concerning an potential of inhibition by the amount of cDNA solution (Chapter 6.2.2.1).

6.2.2.1 Maximum cDNA per slot

For determination of mRNA levels in homogenate, the first step was to guarantee an efficient PCR reaction. The maximum cDNA concentration per one slot was determined at which the PCR was not inhibited. Since the inhibition of a PCR by the amount of cDNA solution is independent of primers and probes, it was sufficient to consider only one transcript, e.g. reference gene (HPRT), for this evaluation.

Different volumes of human mammary gland cDNA per slot were analyzed with the TLDA-WNT32 (Chapter 5.1.6). Since 50 μl cDNA were the maximum load for one slot, 50 μl, 25 μl, 10 μl and 5 μl cDNA were used. For evaluation, the initial copy number (n0) per μl cDNA of each Ct value (triplet) and the relative standard deviations (SDs, percentage of the mean n0) were calculated. If there was no inhibition of the PCR no difference between the n0/μl cDNA of each slot was expected. 50 μl of cDNA resulted in n0 (HPRT) of 7.8±2.9 and 25 μl in a n0 of 73.1±3.6 (Figure 18). This difference was statistically significant ($p<0.0001$, ANOVA, Scheffe) meaning an increase in transcript level with less cDNA which is not possible. Thus, PCR with 50 μl of cDNA was inhibited. Without inhibition of PCR the expected n0/μl cDNA of 10 μl and 5 μl cDNA per slot would be the same. The n0 between 10 μl and 5 μl cDNA per slot showed no significant difference. But there were significant differences between n0/μl cDNA of 25 μl and 10 μl and 5 μl cDNA per slot (25 μl: n0 73.1±3.6, 10 μl:

n0 56.8±6.3, 5 μl: n0 59.1±4.8). But the n0/μl cDNA of 25 μl was higher than expected, in case of an inhibition of the PCR the n0/μl cDNA would be less than expected. In conclusion, for all future experiments equal to or less than 25 μl cDNA per slot were used.

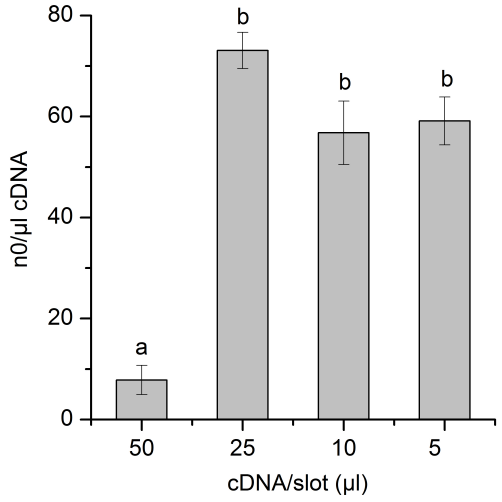


Figure 18: n0-values (mean±SD) of HPRT of three dependent reaction on one TLDA-WNT32 of 5 to 50 μl cDNA per slot. Different letters indicate different data groups. Statistical analysis was performed by means of ANOVA, Scheffe, $p<0.01$.

6.2.2.2 Variations within one slot

For determination of the variations of one triplicate within one slot of TLDA cards, TLDA-WNT32 (32 transcripts, 4 samples per card) and WNT-screen (16 transcripts, 8 samples per card) were used (Chapter 5.1.6). The hypothesis was that the variation in a triplicate is independent of primer and probes and only influenced by distribution of the cDNA on these three spots. Thus, all samples which had been analyzed in the working group using TLDA-WNT32 and TLDA-WNT-screen were taken into account. All samples analyzed were cDNA solutions of Ishikawa cells treated with different estrogens and solvent controls. In sum, 2712 reactions of 30 different transcripts were evaluated (904 means and standard deviation of Ct values). For evaluation of variations of the same cDNA solution analyzed in triplicate PCR reactions on a TLDA, n0 of each Ct value were calculated and the relative SDs (percentage of the mean n0) of each triplicate were represented as the frequency distribution dependent on the absolute mean Ct values. As a first approach, Ct values were divided into ranges from 15-20, 20-25, 25-30 and 30-35 (Figure 19).

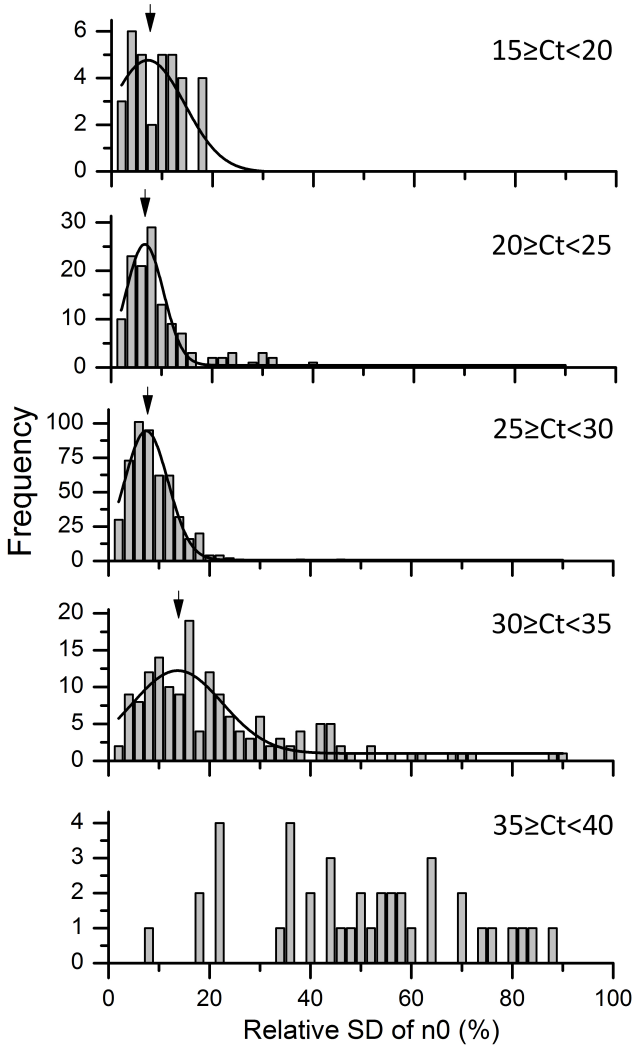


Figure 19: Frequency distribution of 904 relative SDs (percentage of the mean n0) of n0 (triplicate in one slot of TLDA) of 30 different transcripts dependent on their absolute mean Ct value. Mean Ct values were divided in five classes: 15-20, 20-25, 25-30, 30-35 and 35-40. Gaussian fit was performed in all Ct-ranges and modal value, half width and coefficient of determination were evaluated. 15-20: modal value=7.2%, half width=17.2%, R^2=0.786; 20-25: modal value=6.7%, half width=8.7%, R^2=0.936; 25-30: modal value=7.5%, half width=10.2%, R^2=0.966; 30-35: modal value=13.7%, half width=21.0%, R^2=0.740; 35-40: fit failed. The black arrow indicates the modal value.

Ct-range from 10-15 was omitted from this evaluation because of insufficient number of data points. Ct-ranges from 15-20, 20-25 and 25-30 had narrow distributions which was reflected in the half widths of the Gaussian fits of the data 17.2% (R^2=0.786), 8.7% (R^2=0.936), 10.2% (R^2=0.966), respectively. The high coefficients of determination indicated a Gaussain distribution of the realtive SDs. The modal values of the fits were caculated in order to chracterize position of the distribution. The modal values (15-20: 7.2%, 20-25: 6.7%, 25-30: 7.5%) of these three Ct-ranges were very low indicating an avarage SD of one

triplicate in one slot less than 10% within a Ct-range from 15-30.

In contrast, the distribution of the Ct-range between 30-35 was wider (half width: 21.0%) and the modal value was higher (13.7%), but coefficient of determination was still good (R^2=0.740). This indicated a SD >10% of the triplicate in the Ct-range 30-35. For the distribution of the Ct-range from 35-40 the non-linear fit failed and almost all relative SDs were above 30%, which is unacceptably high for quantitive analysis.

In conclusion, there is a clear positive association between the relative SD and Ct values above 30. Ct-ranges from 15-20, 20-25 and 25-30 resulted in low modal values of the relative SDs, ranging from 6.7% to 7.5% and narrow distributions, ranging from 8.7% to 17.2%. Ct values above 35 had unacceptable high relative SDs, thus all Ct values above 35 in a TLDA analysis with only single reaction per transcript and sample were omitted from further analysis. The modal relative SD of the Ct-range 30-35 (13.7%) was twice as high as the modal SD of the lower Ct-range (6.7-7.5%). To investigate whether the higher variations were equally distributed over the whole Ct-range or if it is just a part of the range with high variations, the Ct-range was divided in smaller ranges. To overcome the problem of insufficient number data points with higher Ct values, ranges were divided as follows: Ct-range 30-32, 30-33 and 30-34. For relative SD of these three ranges a Gaussian fit was used to compare the frequency distributions concerning modal value, half width and coefficient of determination (Figure 20).

For the modal relative SD a slight increase was observed with increasing Ct value (30-32: 10.7%, 30-33: 11.9%, 30-34: 13.3%). The modal relative SD of Ct-range 30-34 was almost as high as the modal realtive SD of Ct-range 30-35 (13.7%), whereas the modal realtive SD of the Ct-range 30-33 was lower. Likewise, the half width of the Gaussian fit was increased with increasing Ct value (30-32: 14.2%, 30-33: 17.7%, 30-34: 20.7%) and again, the half width of the Ct-range 30-34 was as high as the half width of Ct-range 30-35 (21.0%). The coefficient of determination of the fit was good for all three Ct-ranges (30-32: R^2=0.808, 30-33: R^2=0.789, 30-34: R^2= 0.808).

In conclusion, the Ct-range 30-34 was comparable in modal relative SD and half width to Ct-range of 30-35, whereas the Ct-range 30-33 was lower in modal relative SD and in half width. This indicates with Ct values higher than 33 the variation of the triplicate is rapidly increasing. Thus the Ct value, above all results gained in TLDA analysis with only single reaction per transcript and sample were omitted, was lowered from Ct 35 to Ct 33. The upper border is similar to (Mengual et al., 2008). They found *precision dropping at around Ct 30-32*, so Ct values >31 were excluded (Mengual et al., 2008).

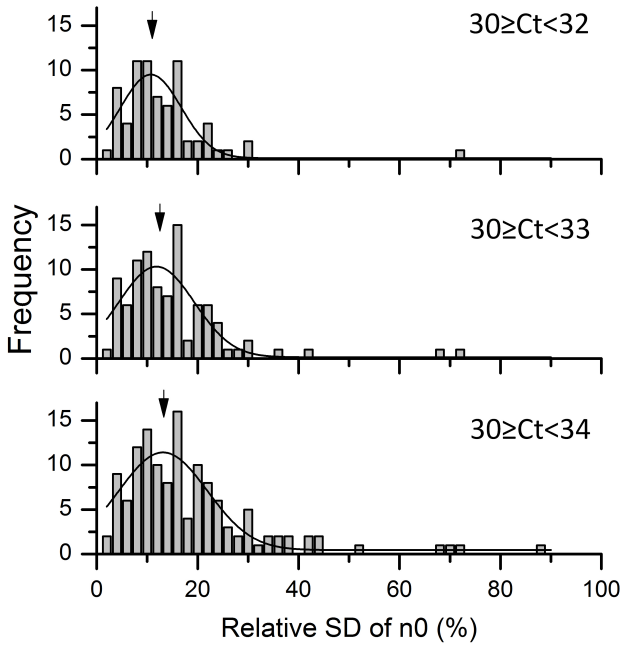


Figure 20: Frequency distribution of 161 relative SDs (percentage of the mean n0) of n0 (triplicate in one slot of TLDA) of 30 different transcripts dependent on their absolute mean Ct value. Mean Ct values were dived in 3 classes: 30-32, 30-33 and 30-34. Gaussian fit was performed in all Ct-ranges and modal value, half width and coefficient of determination was evaluated. 30-32: modal value=10.7%, half width=14.2%, R^2= 0.808; 30-33: modal value=11.9%, half width=17.7%, R^2=0.789; 30-34: modal value=13.2%, half width=20.7%, R^2=0.808. The black arrow indicates the modal value.

Thus, for the customizable format of the TLDA single reaction per transcript could be used with regard to Ct value borders (described above). The format containing 48 transcripts and eight per samples card was chosen, including the internal card control (18S) per sample. Hence, 47 transcripts were put on the TLDA and the remaining two transcripts (UGT1A1 and UGT1A10, chosen randomly) were quantified using TaqMan® probe-based real time PCR in a 96-well format (Chapter 5.2.10).

6.2.3 Variations of one replicate between two slots and two cards

In order to investigate the variations of one replicate in two slots (intra TDLA variation) and on two cards (inter TLDA variation), the TLDA-card chosen with the transcripts chosen (Chapter 6.2.1) was used (TLDA-Polymorphism I, Chapter 5.1.6).

The statistical variation of every transcript (based on n0) was evaluated. Since the amount of RNA from breast tissue homogenate was limited, RNA of MCF7 cells was used. It was expected that transcript levels of MCF-7 and normal human breast tissue differed and were

effected by culture conditions (Lehmann and Wagner, 2008). Therefore, MCF-7 cells were grown in normal culture media or in steroid-free culture media (for 48 h, Chapter 5.2.1). For both culture conditions, two independent RNA isolations were performed and each RNA was reversely transcribed into cDNA (1-4, Figure 21). Then, 13 or 6.5 μl cDNA per slot were analyzed on two TDLA-Polymorphism I cards (Figure 21). Since the absolute n0 of the transcripts may vary from RNA isolation to RNA isolation, the relative SD was calculated from normalized n0 (n0/n0 HPRT, chapter 5.2.10) of each transcript. For intra TLDA variation, the SD of normalized n0 of cDNA2 (cDNA of cells cultured with normal media, TLDA 1, slot 2-4, Figure 21) and cDNA4 (cDNA of cells cultured with steroid-free media, TLDA 1, slot 6-8, Figure 21) were calculated and mean±range/2 of both SDs used for evaluation. Likewise, for inter TLDA variation, the SD of normalized n0 of cDNA1 (cDNA of cells cultured with normal media, TLDA 1-slot 1 and TLDA 2-slot 1-2, Figure 21) and cDNA3 (cDNA of cells cultured with steroid-free media, TLDA 1-slot 5 and TLDA 2-slot 3-4, Figure 21) was calculated and mean±range/2 of both SDs used for evaluation.

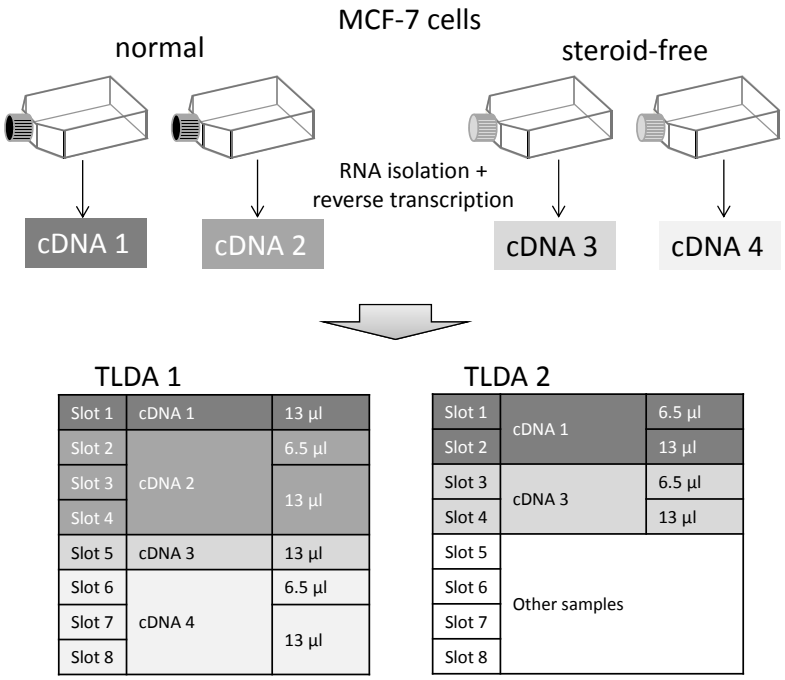


Figure 21: Experimental design of the investigation on intra and inter TLDA statistical variations. MCF-7 cells were grown in normal or steroid-free culture media. For both culture conditions two independent RNA isolations were performed and each RNA was reversely transcribed into cDNA (1-4). Then, 13 or 6.5 μl cDNA per slot were analyzed on two TDLA-Polymorphism I cards. For intra TLDA variation, the SD of normalized n0 of cDNA2 (TLDA 1, slot 2-4) and cDNA4 (TLDA 1, slot 6-8) was calculated. For inter TLDA variation, the SD of normalized n0 of cDNA1 (TLDA 1, slot 1 and TLDA 2, slot 1-2) and cDNA3 (cDNA of cells cultured with steroid-free media, TLDA 1, slot 5 and TLDA 2, slot 3-4) was calculated.

Out of 47 transcripts (mandatory control, 18S, was omitted from analysis due to Ct values <15) 31 transcripts yielded Ct values ≤33 and ≥15 and were evaluated concerning their relative SD. The transcripts of CYP19A1, GUSB, TBP, UGT2B7, UGT1A3/4, UGT1A8, SULT1A1, SULT1E1, SULT2A1, HSD17B1, HSD17B2, GSTP1, ESR2, SFN, NR1I2, NR1I3 were either not detected or yielded Ct values >33 and thus the variations were not analyzed in this experiment.

Intra TLDA variation Ct values determined by triplicate PCR reactions of 14 transcripts varied by <10% (relative SD), of 15 transcripts by <20 and one (SULT1A2) by <30% (Figure 22 upper graph).

The latter was in the Ct-range of 30-33 indicating a Ct value dependent variation (Chapter 6.2.2.2), although Ct values determined for transcripts of MKI67, WNT4 and WNT5A varied by maximum 14.4±1.3%.

Inter TLDA variation Ct values of triplicate PCR performed on two different TLDA cards of 13 transcripts had a mean relative SD ≤ 10%, of 13 transcripts a mean relative SD ≤ 20%, of two transcripts a mean relative SD ≤ 30% and of two transcripts a mean relative SD >30% (Figure 22 lower graph). Two of the three transcripts with a mean relative SD above 20% had a Ct value less than 30 (Ct-range 20-25: TGFB1, Ct-range 25-30: GSTT1, Ct-range 30-33: WNT5A) indicating another factor influencing the inter TLDA variation except for Ct value. Since the intra TLDA variation for these transcripts were not conspicuous and the amplification curves were comparable (same minimum and maximum of the fluoresence intensity and same curve shape) the only reason could be a different threshold. The threshold is automatically set by the SDS software after each TLDA PCR run individually for every transcript (Chapter 5.2.10). For two of the three transcripts the automatically set thresholds were very different between TLDAs (TGFB1: 0.240 and 0.410, WNT5A: 0.494 and 0.368). A 0.2-lower threshold for the same curve resulted in a 1.3 lower Ct value which is less than a half of the n0 of this transcript.

In order to determine the influence of the threshold setting, three different settings were compared:

i Threshold was set to "0.2" for each transcript on every TLDA, manually, hereinafter referred to as "0.2" threshold setting.

ii Threshold was set for each transcript and for each TLDA individually, automatically by SDS software (as earlier described, Chapter 5.2.10), hereinafter referred to as "SDS" threshold setting.

iii Threshold was set for each transcript individually at same value for both TLDA, automatically by ExpressionSuite software (Chapter 5.2.10), hereinafter referred to as

"ExpressionSuite" threshold setting.

With the threshold set to "0.2", TGFB1 and WNT5A had a much lower relative SD for comparison between two TLDAs, 9.0±0.8% (range/2) and 18.9±5.9% (range/2), respectively. However, for the other transcripts either no change or a higher relative SD (ARNT: 7.5±1.7% for automatic threshold to 14.2±4.7% for threshold 0.2) were observed. Additionally, the evaluation of one TLDA with threshold 0.2 resulted in seven transcripts with higher relative SD (SULT1A3/4, PGR, AHR, WNT4, TP53, BMF, ARNT, Figure A57), indicating that the same threshold for all transcripts is not as good as the threshold set by SDS. The relative SD gained with the ExpressionSuite setting were comparable to the results for intra and inter TLDA variation obtained with the threshold setting at 0.2 (Figure A57).

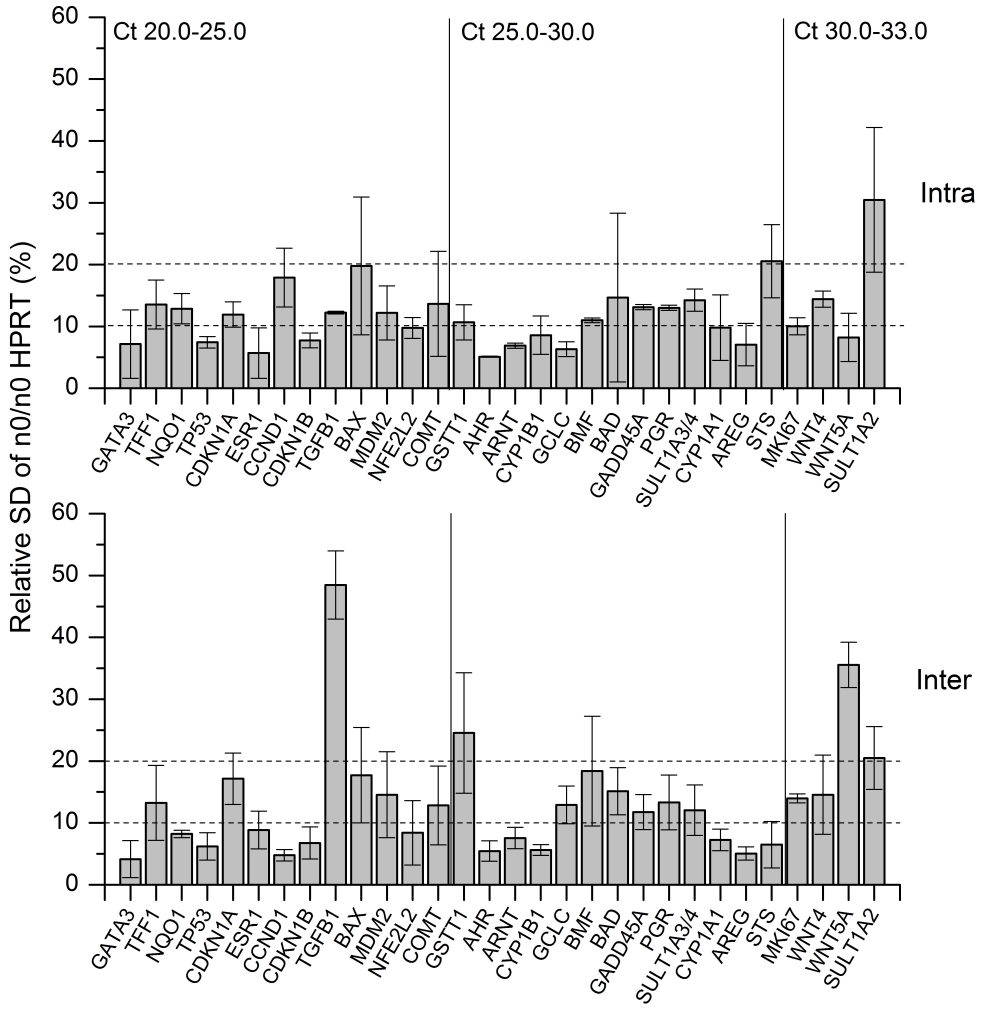


Figure 22: Relative SD of the same sample in three slots on one TLDA (intra variation, upper graph) and two TLDAs (inter variation, lower graph). Data represent mean±range/2 of two independent samples of MCF-7 cells cultured with normal or steroid-free culture media. Out of 47 transcripts 31 transcripts yielded stable Ct values (<33 and >15). 21 of them had SDs <10%. 6 with SD <20% and 3 with SD >20%.

In conclusion, 0.2 and ExpressionSuite threshold setting resulted in higher intra TLDA variation. For intra TLDA variation the SDS threshold yielded the lowest intra TLDA variation. For inter TLDA variation only for Ct values of two transcripts (TGFB1 and WNT5A) lower relative SDs were observed with the 0.2 and ExpressionSuite threshold setting. Thus, the SDS threshold setting seems to generate the lowest variation, but all transcripts of all analyzed TLDA should be monitored concerning their thresholds, since big differences can occur which affect the initial copy number. But this is error-prone and laborious when analyzing more than two TLDAs. Setting the threshold at 0.2 is also difficult when analyzing more than two cards, since the absolute minima and maxima of every amplification curve may vary from PCR run to PCR run. Thus, when when comparing samples measured on

more than two TLDAs the third method (ExpressionSuite) for threshold setting was used.

6.2.4 Linearity of preamplification and pre-preamplification

In order to quantify the mRNA level in human mammary gland homogenate and epithelial cells (laser capture microdissection, Chapter 6.4.1) a TLDA based method had to be developed to determine all transcripts involved in E1/E2 metabolism, marker enzymes for stress, apoptosis, proliferation and cell cycle (Chapter 2.2). Since some transcript levels were expected to be very low in homogenate of normal breast tissue, e.g. isoenzymes of UGT (Lehmann and Wagner, 2008), and only a limited number of epithelial cells were expected to be gained with laser capture microdissection, the method needed to be very sensitive to determine all transcripts in homogenate and epithelial cells.

As starting point for sensitivity optimization, RNA of 200 mg normal human breast tissue homogenate was isolated and reversely transcribed into cDNA and transcripts were quantified on TLDA-PolymorphismI (Chapter 5.2.6 and 5.2.10). Twelve of 47 transcripts yielded in no amplification curve and three had a Ct value>33 (Table 59). The transcript with the lowest Ct value (except for the mandatory card control, 18S) was GATA3 (26.2) and with the highest Ct value CYP19A1 (37.4), the Ct value of the reference gene HPRT was 30.6.

Thus, for detection of hitherto transcripts which gave not rise to an amplification curve until cycle 40 in a quantifiable Ct-range (Ct value<33) in breast tissue homogenate at least 128 times more (seven Ct values less) tissue was needed equaling more than 26 g mammary gland. Since such an amount was impossible to obtain, the possibility of preamplification was evaluated.

To evaluate the requirements for analysis of epithelial cells, 500-1000 epithelial cells (approx. 100 elements) were dissected (Chapter 6.4.1), since this was the maximum number of elements which could be dissected within one day. RNA was isolated (Chapter 5.2.6) and reversely transcribed into cDNA (Chapter 5.2.9). Assuming the mean eukaryotic cell mass of 2.5 ng (Park et al., 2008), 200 mg tissue consists of approx. $8x10^7$ cells analyzed in homogenate, compared to 500-1000 laser dissected epithelial cells. Since with cDNA of homogenate 32 transcripts could be quantified, it was not expected to detect the majority of the 47 transcripts in cDNA of only 500-1000 epithelial cells. Hence, transcript levels of HPRT and cytokeratin 18 (described to be highly expressed in epithelial cells, Bocker et al. 2009) were quantified in cDNA of laser dissected cells using 20 μl-real time PCR (Chapter 5.2.10). However, only the transcript of CK18 was detected (Table 60), while no signal of HPRT was detected up to 40 cycles. Thus, minimum 1024 times more (ten Ct values less) epithelial cells were needed to detect HPRT in a quantifiable Ct-range (15-33, Chapter 6.2). The most transcripts were lower expressed in mammary gland tissue than HPRT (up to seven Ct values more than HPRT in homogenate), therefore $1.3x10^5$ times more (17 Ct

values less) cells would be needed to detect all transcripts, which was impossible for laser capture microdissection.

One commonly used method to increase the copy number of transcripts more than 100 times is a linear amplification prior to real time PCR (preamplification). It is important that the preamplification is in linear range of amplification for all transcripts, because a saturated reaction would change the ratio to the reference gene and thus bias the results.

Analysis of cDNA derived from GECS needed reduction of Ct values by 17. Therefore, one preamplification reaction was not sufficient.

Hence, the preamplified cDNA of epithelial cells was preamplified for a second time (pre-preamplification). Overall, 24 cycles of preamplification were necessary to obtain 17 Ct values less for each transcript, because of two dilutions steps (1:10) due to manufacture's protocol. Moreover, the pre-preamplification reaction was used to investigate the linearity of the preamplification of transcripts which were not quantifiable without preamplification.

In order to investigate linearity of the preamplification reaction, commercially available human mammary gland RNA (Chapter 5.1.5) was reversely transcribed. Since the commercial RNA was more concentrated than the isolated one, it was diluted 1:1000 in ncf-water prior to preamplification to simulate a sample containing a low amount of RNA. The manufacture's protocol recommend 10 to 14 cycles of preamplification which equals 6.7 - 10.7 lower Ct values (104 to 1663 time more n0) for each transcript, due to a 1:10-dilution of cDNA in the preamplification reaction mixture (Chapter 5.2.10). To investigate the influence of the amount of preamplification cycles, the diluted cDNA was preamplified for 10, 12 and 14 cycles using the customized commercially available preamplification pool ("TLDA-PolymorphismII", Figure 23). Then, the preamplified cDNA (14, 12 and 10 cycles) was diluted (1:100) in ncf-water (to simulate low amount of transcripts) and preamplified again for 10, 12 and 14 cycles, respectively (pre-preamplification). Thus, the sum of preamplification cycles was 24 for every combination (Figure 23).

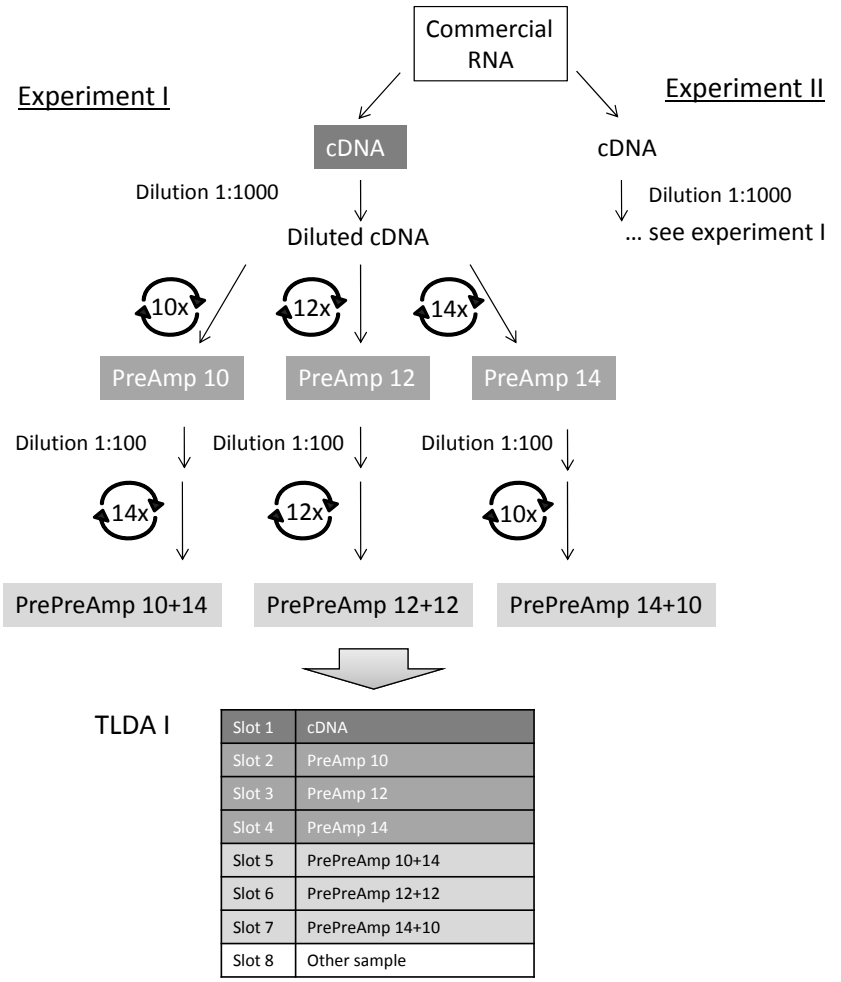


Figure 23: Experimental design of determination of linearity of preamplification and pre-preamplification. Commercially available mammary gland RNA was reversely transcribed into cDNA, diluted (1:1000) in ncf-water and preamplified (PreAmp) for 10, 12 or 14 cycles using preamplification pool-PolymorphismII (Chapter 5.1.6). Then, PreAmps were diluted (1:100) in ncf-water and preamplified for a second time (PrePreAmp) for 10, 12 or 14 cycles. Then transcript levels of cDNA, PreAmp and PrePreAmp were quantified using TLDA-PolymorphismII. The experiment was carried out twice using two independent reversely transcribed cDNAs (experiment I and experiment II).

The whole experiment (preamplification and pre-preamplification) was carried out twice using two independent reversely transcribed cDNAs (Figure 23). The transcript levels were then quantified with the TLDA-PolymorphismII (Chapter 5.1.6). Ct values were calculated using the threshold automatically set by SDS software, since each experiment was carried out on one TLDA (Chapter 6.2.3). For evaluation of linearity of all transcripts Pearson linear regression of (non) preamplified against (pre-)preamplified Ct values (previously described for preamplification in Khan et al., 2012; Asztalos et al., 2010) were used (Chapters 6.2.4.1 and 6.2.4.2).

Transcripts, which were not included in the TLDA-PolymorphismII (UGT1A1 and UGT1A10),

were quantified in a separate PCR 20 μl reaction using TaqMan® Assay (Chapter 5.2.10). Since only two transcripts were not enough for linear regression- or box plot analysis, the resulting ΔCt values of (non) preamplified Ct values minus (pre-)preamplified Ct values (method previously described in Noutsias et al., 2008; Li et al., 2008) were compared to the expected ΔCt values (Chapter A9.3.1).

6.2.4.1 Preamplification

All Ct values of transcripts quantifiable without preamplification (Ct values 15-33, Chapter 6.2.2.2) were plotted against their corresponding Ct values of preamplified cDNA (14, 12 and 10 preamplification cycles). Ct values greater than 33 were detected for transcripts of CYP19A1, UGT1A3/4 UGT1A8, SULT2A1, SULT1E1 and NR1I3. SFN was not detected without preamplification and the mandatory card control (18S) was omitted in preamplification reaction, since the high amount of mRNA of 18S may bias the preamplification reaction.

The resulting coefficients of determination (Pearson linear regression) for 10 (R^2=0.91), 12 (R^2=0.89) and 14 (R^2=0.91) preamplification cycles (Figure 24A) were lower than expected (R^2=0.95 Khan et al. 2012 to R^2=0.98 Asztalos et al. 2010). Four data points (in "10 and 14 preamplification cycles") and six data points ("12 preamplification cycles") had larger distances to the fit than the others (Figure 24A). These data points were identified as Ct values of BAD, MDM2, MKI67, GADD45A, GSTP1 (in "10 and 12 preamplification cycles"), NQO1 (in "12 and 14 preamplification cycles") and SULT1A3/4 (in "12 preamplification cycles"). In order to statistically identify outliers, Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007) was applied: Every data point with more than 1.5x inter quartile range to the median were identified as outliers which correspond to the whisker of the boxes in the box plots presented in this work.

With box plot of ΔCt values (non preamplified Ct values - preamplified Ct values, Figure A58) the mentioned transcripts were identified as outliers and considered not preamplifiable and excluded from further analysis. The coefficients of determination of the linear regression of the Ct values of the remaining transcripts (0.98, 0.98 and 0.97 for 10, 12 and 14 cycles, respectively, Figure A24B) were in the expected range of 0.95 (Khan et al., 2012) to 0.98 (Asztalos et al., 2010).

Likewise, in experiment II (Figure 23) the coefficients of determination were 0.87, 0.84 and 0.79 for 10, 12 and 14 cycles of preamplification, respectively (Figure A59A). With box plot, the ΔCt values of MDM2, BAD, GSTP1, NQO1, AHR, GADD45A and HSD17B1 (in "10 and 12 preamplification cycles"), CYP1A1 and COMT (in "14 preamplification cycles") were identified as outliers (Figure A60). After excluding outliers, the coefficients of determination

were (0.98, 0.98 and 0.97 for 10, 12 and 14 cycles, respectively, Figure A59B) in the expected range.

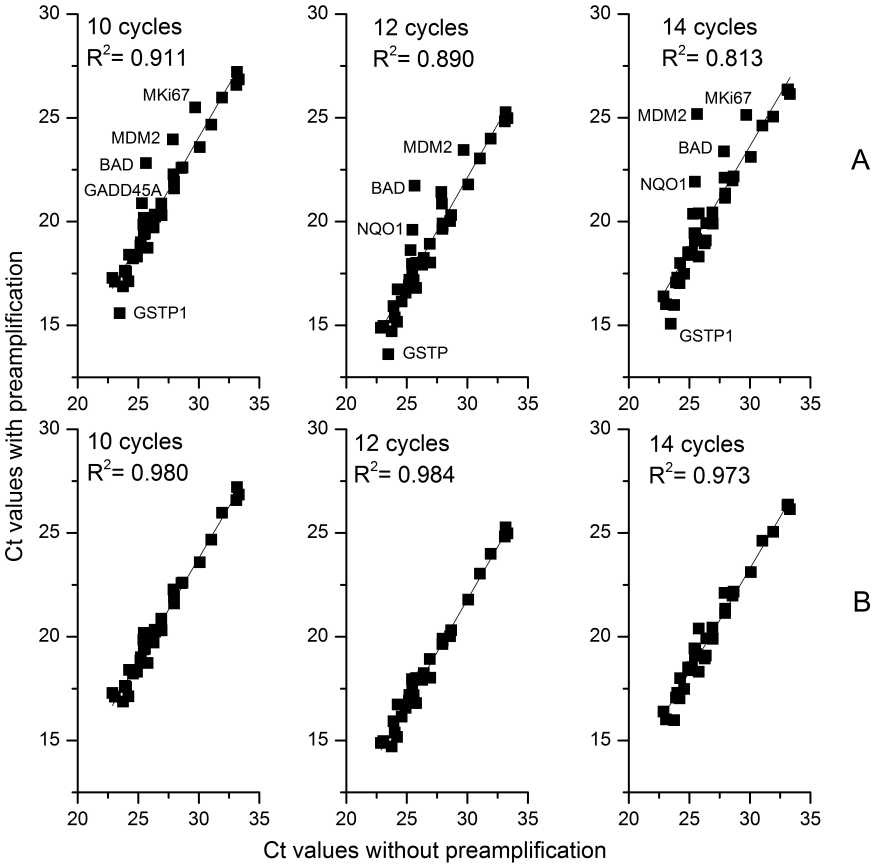


Figure 24: Preamplification curves (Ct values of non preamplified cDNA vs. preamplified cDNA) of experiment I with 10, 12 and 14 cycles of preamplification. Commercial available RNA was reversely transcribed into cDNA and preamplified. Then real time PCR with TLDA-Polymorphisms II with and without preamplification was performed. (A) Ct values of all quantifiable transcripts. (B) Ct values of quantifiable transcripts excluding outliers. Outliers were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

In conclusion, for some of the tested transcripts the preamplification reaction seemed to be not linear (summarized in Table 40). MDM2, BAD and GSTP1 were not preamplifiable in both experiments and in all tested preamplification cycles. GADD45A and NQO1 were not preamplifiable in all but one reactions.

Table 40: Non preamplifiable (PA) transcripts in two independent experiments (I and II), Figure 23.

Considered as	Transcripts					
	10 cycles		12 cycles		14 cycles	
	I	II	I	II	I	II
Non PA all reactions	MDM2 BAD GSTP1	MDM2 BAD GSTP1	MDM2 BAD GSTP1	MDM2 BAD GSTP1	MDM2 BAD GSTP1	MDM2 BAD GSTP1
Non PA most reactions	GADD45A	GADD45A NQO1	GADD45A NQO1	GADD45A NQO1	GADD45A NQO1	NQO1
PA	MKI67	AHR HSD17B1	MKI67 SULT1A3/4	AHR HSD17B1	MKI67	CYP1A1 COMT

Coefficients of variation resulting from preamplification reactions using 10 and 12 cycles were higher than that resulting from 14 preamplification cycles (Figure 23). Thus, 14 preamplification cycles affected transcript levels more than 10 and 12 preamplification cycles,

although the coefficients of determination were in the range of previously described experiments (Khan et al., 2012; Asztalos et al., 2010). For this work 12 cycles of preamplification were used for quantification of transcript level in human mammary gland homogenate, because it is comparable to 10 cycles in coefficients of determination and two Ct values less could be obtained for quantification of less abundant transcripts.

Preamplification of NQO1 and GSTP1

The evaluation of the absolute transcript levels of transcripts with biased preamplification was not possible. However, in preamplification reactions with commercial human mammary gland RNA the transcript levels of NQO1 were underestimated and GSTP1 overestimated (Figure 24). Hence, the hypothesis was, that the direction and amount of bias of preamplification on transcript levels maybe comparable for one transcript in different samples. Thus, a relative evaluation of the preamplified transcript levels among the samples would be applicable.

To collect more data concerning the effect of preamplification on both transcript levels, the transcript levels of NQO1 and GSTP1 were quantified in cDNA in addition to the quantification in preamplified cDNA for mammary gland tissue homogenate of 25 samples. However, neither the direction (over- or underestimation) of the preamplification bias nor the relative amount of bias were comparable in all samples (Figure 25). With preamplification, the transcript level of NQO1 was overestimated in sample #20 while it was underestimated in all other samples. The transcript levels of GSTP1 were overestimated in seven samples and underestimated in 18 samples. Since the direction and the amount of the bias of the preamplification reaction on both transcript levels were not comparable among the samples,

it is likely that the bias in preamplification of other non preamplifiable transcripts (BAD, MDM2, GADD45A) was also not comparable among all samples. Therefore, a quantification of these transcripts without preamplification in all samples was necessary.

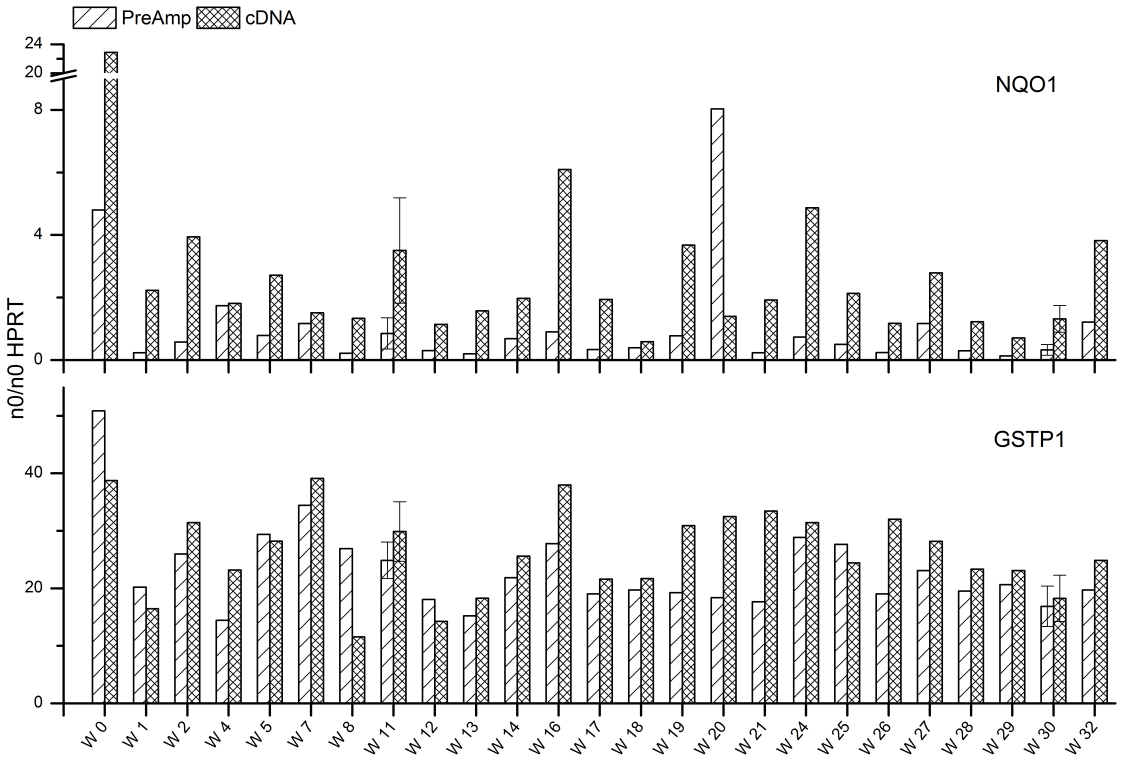


Figure 25: Comparison of transcript levels of NQO1 and GSTP1 in mammary gland homogenate with preamplification (PreAmp) and without PreAmp (cDNA) of 25 samples (w). Preamplified cDNAs (12 cycles) were analyzed with TLDA-PolymorphismII and non premaplified cDNAs in 20 μl real time PCR both in single reaction of one tissue aliquot of each sample, except for samples #11 and #30. Both transcripts were quantified in six different tissue aliquots of sample #11 and #30. For both the mean$\pm$SD is presented.

Since GSTP1 and NQO1 play a major role in E1 and E2 metabolism both transcripts were quantified without preamplification in a separate real time PCR (Chapter 5.2.10). BAD, MDM2 and GADD45A were not further analyzed in this work. MKI67, AHR and HSD17B1 were not preamplifiable in either experiment I or experiment II, indicating a rather random effect. Likewise, SULT1A3/4, CYP1A1 and COMT were not preamplifiable only in one preamplification of one experiment. Hence, the six transcripts were considered as preamplifiable and not excluded from further analysis.

6.2.4.2 Pre-preamplification

In order to investigate the linearity of the preamplification reaction of transcripts which

were not in the quantifiable Ct value range without preamplification (CYP19A1, UGT1A3/4 UGT1A8, SULT2A1, SULT1E1 and NR1I3) and quantification of transcripts of epithelial cells, a pre- preamplification was performed (Figure 23). The preamplified cDNA was diluted (1:100) prior to pre-preamplification, because the Ct value of HPRT of preamplified cDNA of commercial RNA was too low (approx. Ct value = 17) for a additional preamplification reaction.

All transcripts in quantifiable Ct value range (Chapter 6.2.2.2) with 14, 12 and 10 cycles of preamplification were plotted against the corresponding Ct values with (14+)10, (12+)12 and (10+)14 cycles of pre-preamplification. SFN and 18S (mandatory card control) were again excluded.

No literature data were available on the coefficients of correlation of pre-preamplified Ct values against preamplified Ct values. However, the comparison of preamplified with pre-preamplified Ct values and non-preamplified with preamplified Ct values is similar, since only one preamplification reaction is considered. Thus, the same range of the expected coefficients of determination (0.95 to 0.98 Khan et al., 2012; Asztalos et al., 2010) was applied. The coefficients of determination of the linear regression of Ct values in both experiments (I and II) of 10 (R^2=0.95 and 0.97), 12 (R^2=0.92 and 0.81) and 14 (R^2=0.92 and 0.81) pre-preamplification cycles were not in (or at lower limit of) the expected range (Figures A61 and A63). With the box plot, the same outliers were identified (Figures A62 and A64) as in preamplification experiments (Chapter 6.2.4.1) in almost all reactions (MDM2, BAD, NQO1 and GADD45a, Table 41). Comparable to preamplification experiments, some transcripts (SULT1A3/4, SULT2A1, CYP1A1, UGT1A3/4, UGT1A8 and NR1I3) were outliers in only 1-3 reactions (Table 41), indicating a rather random effect. After excluding outliers the coefficients of determination were 0.99 (0.99), 0.99 (0,99) and 0.98 (0.97) for 10, 12 and 14 cycles of experiment I (experiment II), respectively (Figures 61 and 63) and thus in the expected range.

Comparable to preamplification reaction, the coefficients of determination observed after 10 and 12 cycles of pre-preamplification were higher than the coefficients after 14 cycles of pre-preamplification (Figures A61 and A63) with and without exclusion of outliers. Since the pre-preamplification was used for quantification of transcript levels in laser dissected epithelial cells of the mammary gland, 14+10 pre-preamplification cycles were used for practical reasons, despite slightly higher bias in quantification (R^2=0.97 and 0.96, Figures A65 and A67) compared to 12+12 cycles of pre-preamplification. This enables the determination of the quality of the RNA of the dissected cells with only one preamplification reaction, by quantifying the reference gene HPRT. Since, with 14 cycles of preamplification HPRT was already in a quantifiable Ct value range.

Table 41: Non pre-preamplifiable (pPA) transcripts in two independent (I and II) experiments, Figure 23.

Considered as	Transcripts					
	(14+)10 cycles		(12+)12 cycles		(10+)14 cycles	
	I	II	I	II	I	II
Non pPA all reactions	MDM2 BAD NQO1	MDM2 BAD NQO1	MDM2 BAD NQO1	MDM2 BAD NQO1	MDM2 BAD NQO1	MDM2 BAD NQO1
Non pPA most reactions	GADD45A	GADD45A	GADD45A	GADD45A	GADD45A	
pPA	UGT1A3/4 UGT1A8 SULT1A3/4 SULT2A1	UGT1A3/4 CYP1A1		UGT1A8	UGT1A8 NR1I3	UGT1A3/4 SULT1A3/4

Summary of the established method for quantifying 47 transcripts in human mammary gland tissue using TLDA-Polymorphism II:

Maximum cDNA per slot:	25 μl
Lower limit of Ct value for quantification:	15
Upper limit of Ct value for quantification:	33
Method of threshold setting for experiments with 1 or 2 TLDA:	SDS
- Intra TLDA relative SD of all transcripts (mean±SD):	11.9±5.3%
- Inter TLDA relative SD of all transcripts (mean±SD):	13.4±9.5%
Method of threshold setting for experiments with more than 2 TLDAs:	ExpressionSuite
- Intra TLDA relative SD of all transcripts (mean±SD):	17.8±8.7%
- Inter TLDA relative SD of all transcripts (mean±SD):	13.3±5.4%
Cycles of preamplification:	12
Cycles of pre-preamplification:	14+10

Without preamplification, eight transcripts were not in the quantifiable Ct value range (CYP1A1, CYP19A1, UGT1A3/4, UGT2B7, UGT1A8, SULT2A1, SULT1E1, NR1I3, Table 42). All eight transcripts were quantifiable with a preamplification of 10, 12, or 14 cycles. Five transcripts were not preamplifiable: BAD, GADD45A, MDM2, GSTP1 and NQO1. When removing only the five transcripts from (pre-)preamplification curves the coefficients of determination (Range: 0.94-0.99, Figures A69 and Figures A70) were barely in range of previously published studies (0.95 to 0.98 Khan et al., 2012; Asztalos et al., 2010). GSTP1 and NQO1 play an important role in the E1/E2 metabolism. Since the work is focusing on E1/E2 metabolism, both transcripts were quantified in an additional real time PCR with cDNA as template to circumvent the preamplification reaction of both transcripts. The transcripts of BAD, GADD45A and MDM2 need to be quantified in a PCR without preamplification. In this work they were not further analyzed.

Table 42: Transcripts with and without preamplification.

Non quantifiable transcripts without preamplification	Non preamplifiable transcripts
CYP1A1, CYP19A1, UGT1A3/4, UGT2B7, UGT1A8, SULT2A1, SULT1E1, NR1I3	BAD, GADD45A, GSTP1, NQO1, MDM2

6.2.4.3 Preamplification for TaqMan® Assay

The aim was to quantify 49 transcripts. Since the format of the TLDA was restricted to 47, two transcripts (UGT1A1 and UGT1A10) were quantified with TaqMan® Assay in addition to TLDA (Chapter 6.2.1). Since both UGT transcripts were expected to be less abundant in normal breast tissue, the linearity of the preamplification reaction was investigated. However, linear regression and box plot analysis of only two Ct values would not be meaningful. Therefore, ΔCt values of non preamplified and preamplified Ct values (Noutsias et al., 2008; Li et al., 2008) were compared to the expected ΔCt values (calculation in Chapter A9.3.3).

This experiment was performed using commercially available mammary gland RNA which was reversely transcribed into cDNA. The cDNA was preamplified for 10, 12 and 14 cycles. Then, Ct values were determined by means of TaqMan® Assay-based real time PCR (Chapter 5.2.10).

Each preamplified cDNA was diluted so that the expected ΔCt value was 4.7. Then, the percentage of the determined ΔCt value from the expected ΔCt value was calculated and compared to the minimum and maximum ΔCt value for an unbiased preamplification reaction as used for transcripts determined with TLDA (minimum 86% and maximum 104%, Figure A73). The ΔCt values of the transcripts of HPRT and UGT1A1 were in this range. Only the relative ΔCt value of the 12-cycle-preamplification of UGT1A10 (85.4%) was less than but close to the minimum ΔCt value. However the relative ΔCt values for the 10- and 14-cycle-preamplification were both in the range which indicated a random effect of the 12-cycle-preamplification reaction. In conclusion, both transcripts were preamplifiable.

Table 43: ΔCt values relative to the expected ΔCt value ($\%Ct_{exp}$) of preamplified cDNA for 10, 12 and 14 cycles of preamplification (PA) of HPRT, UGT1A1 and UGT1A10.

Transcript	$\%Ct_{exp}$		
	PA10	PA12	PA14
HPRT	102.6	96.0	95.2
UGT1A1	96.6	90.9	90.7
UGT1A10	100.1	85.4	98.4

6.2.5 Influence of sample-taking sites on transcript levels

In order to ensure a correct and comparable quantification of each transcript for each sample, the variation within one sample was investigated.

Since the intra and inter TLDA variations were characterized (Chapter 6.2.3), another factor which may influence the variation of the transcript levels was the individual tissue composition. Different tissue aliquots prepared from one sample may differ in amount of epithelial cells, stromal cells and fat cells (Chapter 6.1), with each cell type having different transcript levels of enzymes investigated in this work (Chapters 2.1 and 2.2).

In order to compare statistical sampling of tissue of breast sides and sites, homogenate of three different tissue aliquots (Chapter 5.2.2) of the left and right breast of two female donors were analyzed (Chapter 5.2.10). RNA of three different tissue aliquots of each breast side were isolated, reversely transcribed, preamplified (12 cycles) and the transcript levels determined using TLDA-PolymorphismII (Chapter 5.2.10). The n0 of the target transcripts were normalized to n0 of the reference gene HPRT for quantitative comparison.

6.2.5.1 Sample-taking sides

For comparison of the sampling of left and right breast of the female donors the mean±SD of each breast side (three aliquots each) of two female donors were evaluated. As expected, no significant differences between left and right breast were observed (Figures A71 and A72) using t-test with subsequent p value adjustment (Holm, Chapter 5.2.15).

6.2.5.2 Sample-taking sites

In order to compare sample-taking sites the six aliquots of each tissue donation were averaged and the relative SD of the transcripts of both donations used for evaluation.

The mean relative SD of all transcripts of two tissue donations (six aliquots each) was 41.7±25.3% (SD±range/2, Figure 26), which was more than three times as high as the mean inter TLDA variation of all transcripts (13.3±5.4%, Chapter 6.2.3). Additionally, the SD of the mean relative SD was more than five times as high compared to the inter TLDA SD of all transcripts.

Huge differences in relative SD of different transcripts were observed, ranging from 15.7±5.0% (AHR) to 108.4±18.9% (CYP1A1, Figure 26). The Ct values of the transcripts of CYP1A1, TFF1, UGT2B7 and SULT1E1 had high relative SD in both homogenates (108.4±18.9%, 103.6±15.2%, 85.6±14.1%, 80.8±6.8%, respectively, Figure 26). With exception of TFF1, these were all very less abundant transcripts (Tables A62 and 63). The Ct

values of NR1I3 and CYP19A1 had only a high relative SD in homogenate of sample #11 (152.2% and 150.6%, respectively, Figure 26).

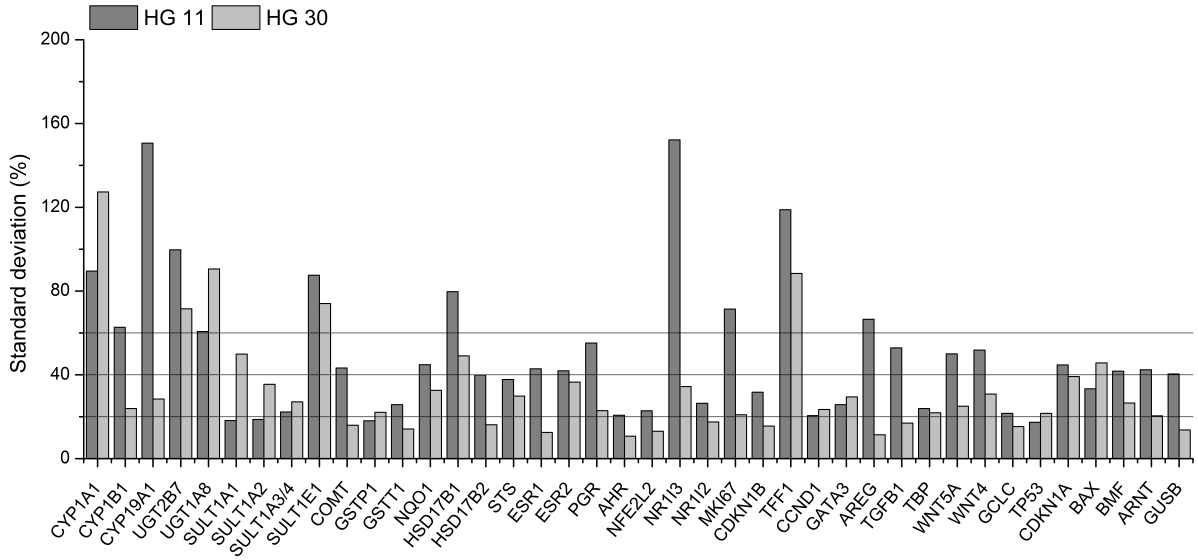


Figure 26: Relative SD of all transcript levels determined with TLDA-Polymorphism II out of RNA of six different tissue aliquots of breast tissue homogenate (HG) of woman #11 and #30.

6.2.5.3 Discussion

In conclusion, huge range in relative SD among transcripts of approx. 93% was observed.

To assess the relevance of the variation of the transcript levels of one sample for the further analysis, the inter individual variation was taken into account. The transcript levels with the highest variations in samples #11 and #30 (CYP1A1, CYP1B1, CYP19A1, UGT2B7, UGT1A8, SULT1E1, HSD17B1, NR1I3, MKI67, TFF1 and AREG) were compared to the respective transcript levels of the remaining 23 samples (Figure 27). Additionally, transcripts with high inter-sample variation (CYP1B1, TFF1, AREG) are only expressed in epithelial cells of the mammary gland tissue (Chapters 2.2.2 and 2.1), indicating the possible role of different tissue composition among the aliquots of one sample for inter-sample variation.

The range of the relative SD of these transcripts of tissue from both donors was never higher than range of the respective transcript level of all 25 samples, indicating higher inter-individual variation than inter-sample variation. Thus, differences in the individual transcript levels were still detectable and thus the results can be used for further statistical evaluation. However, for the mentioned transcripts, the result should be critically reviewed and challenged concerning plausibility.

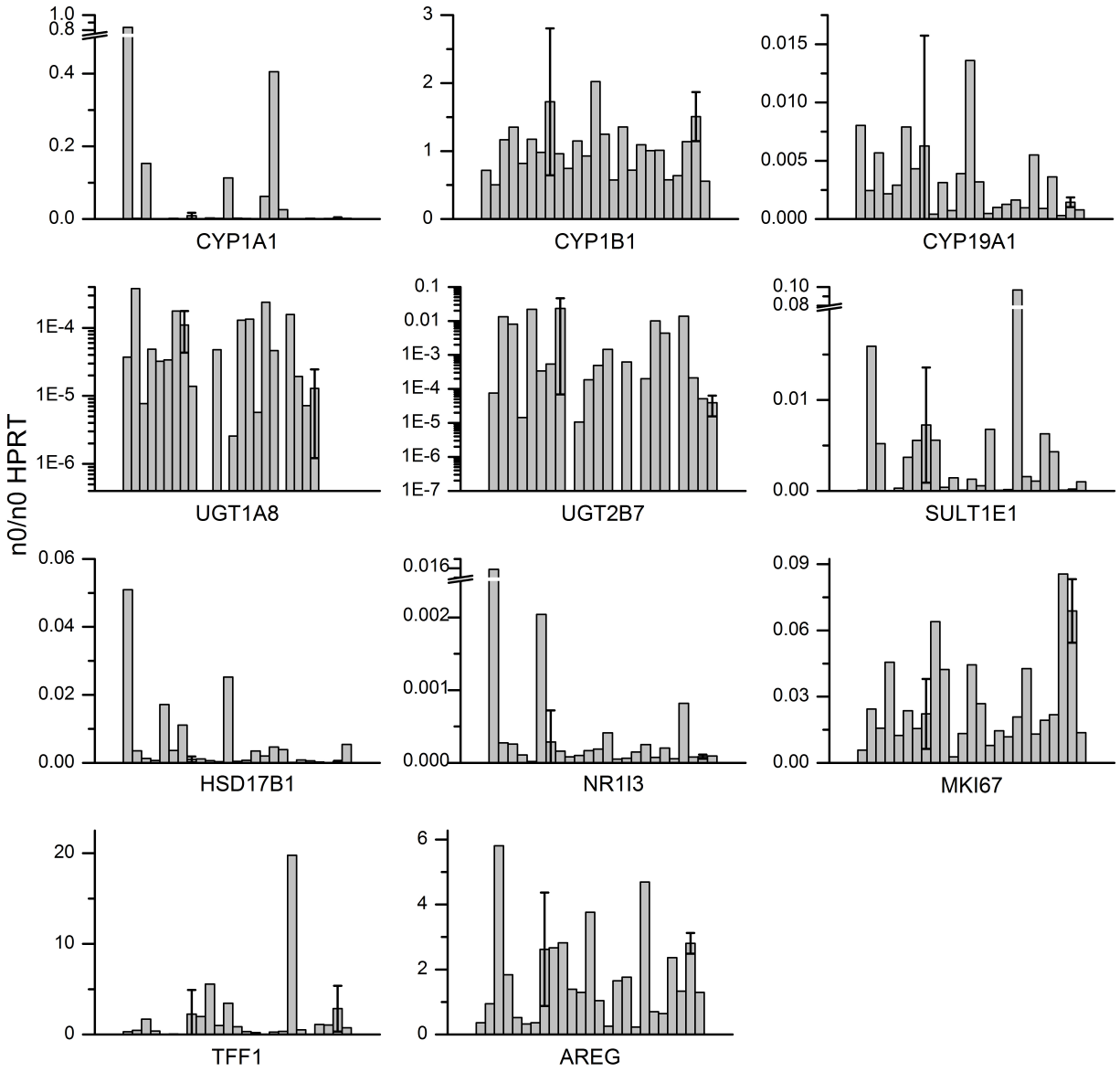


Figure 27: Transcript levels of homogenate (n0/n0 HPRT) with the highest intra sample variation of samples #11 and #30. Presented are the values of the single determination of 23 samples in in comparison to the mean±SD of six aliquots of samples #11 and #30. Transcript levels were determined using TLDA-PolymorphismII and analyzed with SDS software (Chapter 5.2.10).

6.3 Transcript levels in mammary gland tissue homogenate

For some of the transcripts investigated in this work, data available on presence or quantitative levels in normal mammary gland tissue were insufficient or nonexistent (e.g. UGTs, Chapter 2.2.2). Furthermore, differences in enzymes and receptors of normal tissue compared to MCF-7 cells, a wildly used breast cancer cell line for studying factors influencing breast cancer risk, are described (e.g. Lehmann and Wagner, 2008), but mostly qualitatively and not for all important transcripts involved in E2 metabolism (Chapter 2.2).

Therefore, the transcript levels of the normal human mammary gland tissue samples were quantified, compared to literature data available and additionally compared to transcript levels determined in MCF-7 cells.

Five out of 30 mammary gland tissue samples did not yield enough RNA for quantitative analysis of less abundant transcripts (0.00001-fold of HPRT or less, Table A61).

Thus, 25 samples were analyzed concerning their transcript levels of genes involved in E1/E2 metabolism, marker enzymes for stress, apoptosis, proliferation and cell cycle (Chapter 6.2.1).

6.3.1 Qualitative differences of transcript levels

Of the 49 transcripts analyzed in this work eight were not determined (Ct values >33, Chapter 6.2.2, or no detectable increase in fluorescence signals until cycle 40) in every sample: The level of UGT1A3/4 was not determined in any sample (Figure 28). Furthermore, none of the other isoenzymes of UGT family were determined in every sample (Chapter 2.2). The transcript of UGT1A1 was determined in 14 of 25 samples, UGT1A8 in 21, UGT1A10 in 22 and UGT2B7 in 21 of the 25 samples.

In accordance, the mRNA of UGT1A8 and UGT2B7 were previously detected in very low amounts in mammary gland RNA (Chapter2.2.2). The data on the expression of UGT1A10 and UGT1A3/4 were inconsistently, since both were detected in one of two studies (Chapter 2.2.2). The transcript of UGT1A1 was not detected in three previously published studies (Chapter 2.2.2). However, all of the previously published results were based on one or two samples. Thus, new findings on the presence of the transcripts of UGT1A1 and UGT1A10 in normal mammary gland tissue were obtained due to a larger number of samples.

All transcripts of the SULT isoenzyme family, except for SULT2A1, were determined in each sample (Figure 28). The transcript of SULT2A1 was detected in only three of the 25 samples.

The presence in the mammary gland tissue of transcripts of the SULT family was previously described (Chapter 2.2.2), except for the transcript of SULT2A1 which was not detected in mammary gland RNA in a previously published study analyzing one sample

(Chapter2.2.2).

Furthermore, the transcripts of SFN, a negative regulator of the cell cycle and sensitive to DNA damage (Chapter 2.3), and CCND1, a marker for proliferation and ESR1 activation (Chapter 2.1), were detected in 18 and 20 of the 25 samples, respectively. The remaining transcripts were detected in each sample investigated in this work (Figure 29).

The transcript level of SFN was very low which is in accordance to previous studies, since it has been described to be expressed only in myoepithelial cells, a minor cell type in mammary gland homogenate (Chapter 2.1).

In contrast, the complete loss of a signal of CCND1 was unlikely due to transcriptional regulation, because this would equal a reduction to 0.0001% of the transcript level (2.4 n0/n0 HPRT vs. no fluorescence signal until cycle 40). Additionally, no CNP has been described for this gene. However, a single nucleotide polymorphism was described (G to A at mRNA position 870, rs9344), which does not result in an amino acid chain change, but interfered with splicing from exon 4 to exon 5, since it is located in a conserved splice region (Betticher et al., 1995). In turn, this could interfere with the TaqMan® assay used (CCND1-Hs99999004_m1, Table 6) which is located at the boundary of exon 4 to exon 5, the exact position of the primers and probe are not stated by Life Technologies.

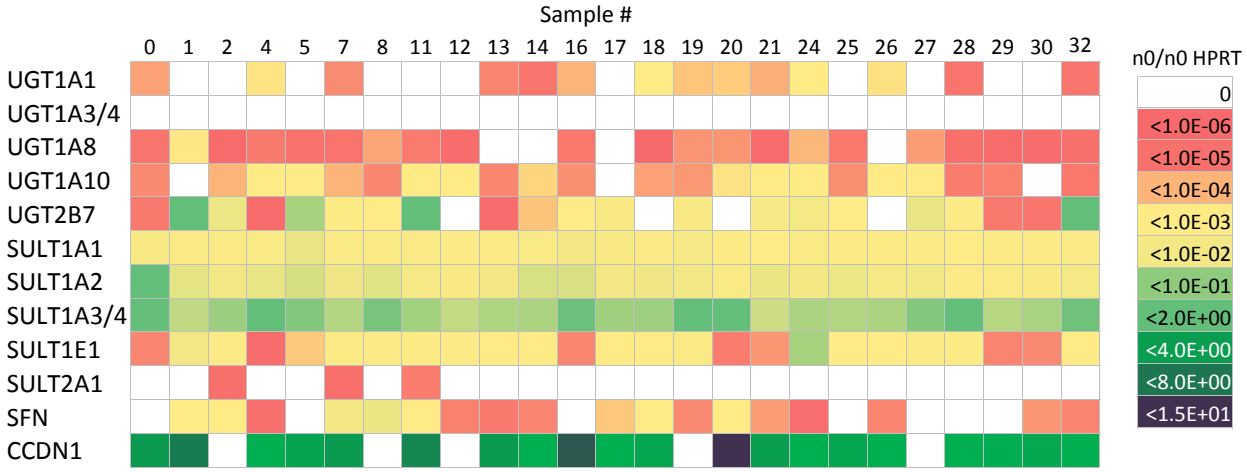


Figure 28: Heatmap of levels (n0/n0 HPRT) of transcripts not determined in all samples (rows) in all mammary gland tissue samples (columns). Levels of transcripts with Ct value <33 or no detectable fluorescence signal were indicated as 0.

6.3.2 Comparison of transcript levels of mammary gland tissue to MCF-7 cells

In order to describe the differences in E2 metabolism of normal mammary gland tissue and MCF-7 cells, a widely used cell line for the investigation of factors influencing E2 metabolism and E2 mediated proliferation, the transcript levels of normal mammary gland tissue were

compared to transcript levels of MCF-7 cells. Cells were cultured with either normal or steroid-free culture media for 48 h, for consideration of both growth stimulated and unstimulated cells, respectively. To ensure comparability among the transcript levels, the Ct values were evaluated with ExpressioSuite software, using the same threshold for every sample (Chapter 6.2.3). In order to compare transcript levels quantitatively, n0 of the target transcripts were normalized to the n0 of the reference gene HPRT (n0 target transcript/n0 HPRT). It must be considered, that the cDNA of MCF-7 was not preamplified, therefore the limit of determination was 20-fold higher for transcripts of MCF-7 cells (approx. $7x10^{-4}$ n0/n0 HPRT) cells than for normal tissue (approx. $3x10^{-5}$ n0/n0 HPRT), based on the Ct value of the reference gene HPRT. Assuming cycle 38 as last cycle where an increase of the fluorescence signal could be detected reliably, the limits of detection in MCF-7 cells were $2x10^{-5}$ n0/n0 HPRT and in normal mammary gland tissue, due to preamplification, $1x10^{-6}$ n0/n0 HPRT.

The levels of BAD, MDM2 and GADD45A were not taken into account for quantitatively comparison, since the quantification of the transcript levels in mammary gland were biased by preamplification (Chapter 6.2.4.1).

Apoptosis

The transcript levels of BMF in normal tissue (median: 0.12 n0/n0 HPRT) were comparable to transcript levels in MCF-7 cells cultured with steroid-free culture media (0.05 n0/n0 HPRT, Figure 29), while the transcript levels of BMF in MCF-7 cells cultured with normal culture media were approx. ten times lower (0.003 n0/n0 HPRT). The transcript levels of BAX were higher in normal tissue (median: 1.2 n0/n0 HPRT) compared to levels in MCF-7 cells of both culture conditions (DMEM: 0.14 n0/n0 HPRT, SF: 0.33 n0/n0 HPRT, Figure 29).

Cell cycle

CDKN1A negatively modulates cell cycle progression (Chapter 2.1), the highest transcript levels were determined in MCF-7 cells cultured with steroid-free media (1.6 n0/n0 HPRT), while the transcript level of CDKN1A was determined in similar range in normal mammary gland tissue (0.5 n0/n0 HPRT) and cells cultured with normal culture media (0.4 n0/n0 HPRT, Figure 29).

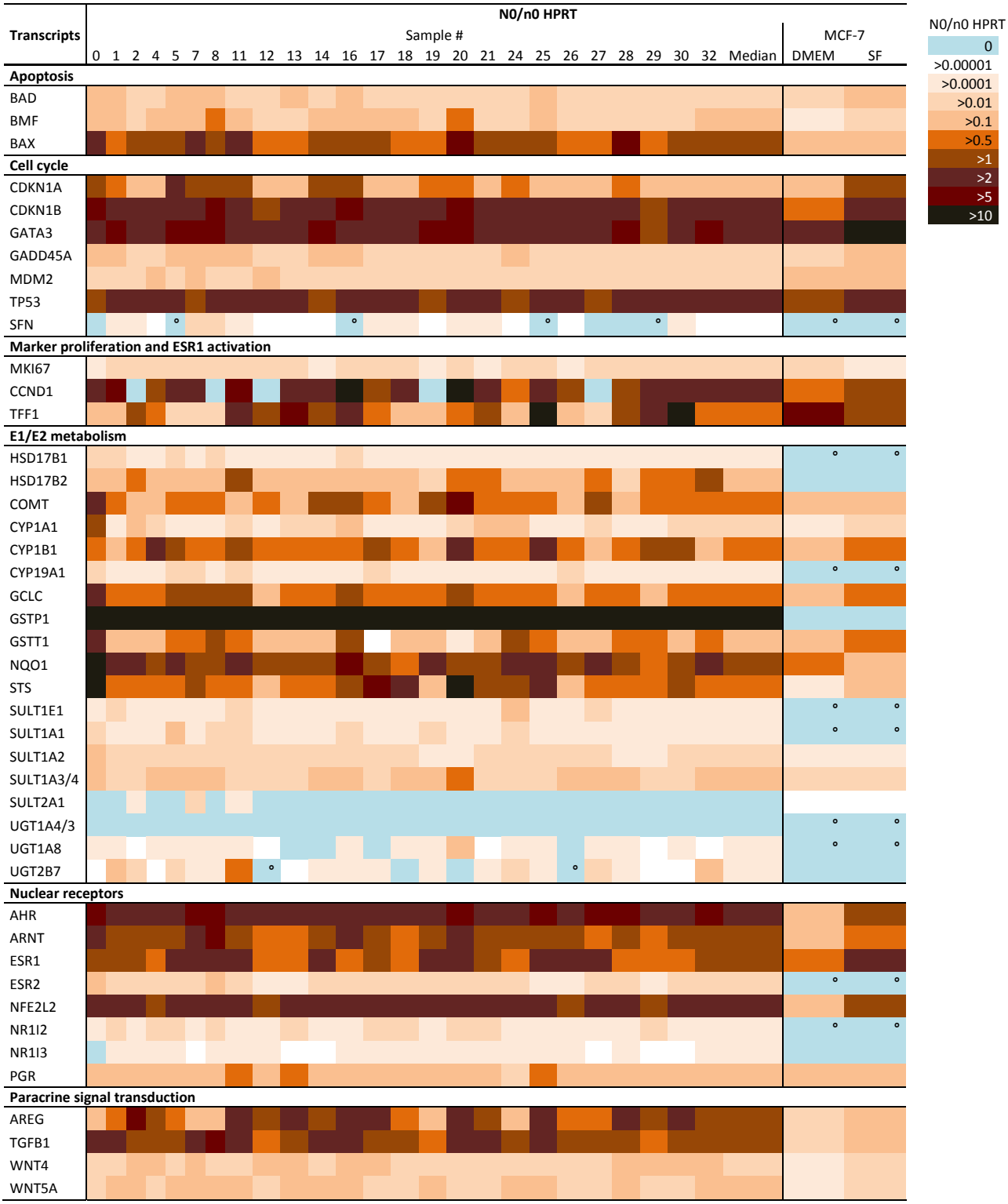


Figure 29: Comparison of transcript levels between normal human mammary gland tissue and MCF-7 cells cultured with normal culture media (DMEM, mean of n=2) and steroid-free culture media (SF, mean n=2). All Ct values were evaluated with ExpressionSuite software. °, fluorescence signal was detected but the Ct value was >33.

Since normal mammary gland tissue of non-pregnant adults mostly consists of senescent cells and the MCF-7 cells cultured in steroid-free media hardly proliferate, a higher level of transcripts, negatively modulating cell cycle, were expected. The transcripts of TP53 and CDKN1B, also involved in negative cell cycle progression (Chapter 2.1), had similar levels in mammary gland tissue and MCF-7 cells cultured with steroid-free media (Figure 29), while the levels of both transcripts were lower in MCF-7 cells cultured with normal media (Figure 29). Since the normal mammary gland tissue was gained from non-pregnant, adult donors, substantial ductal side branching is unlikely (Chapter 2.1). The transcript levels of GATA3, involved in ductal side branching (Chapter 2.1), were comparable in normal mammary gland tissue and MCF-7 cells cultured with normal or steroid-free media (3.3 n0/n0 HPRT, 5.0 n0/n0 HPRT and 10.0 n0/n0 HPRT, respectively, Figure 29).

The transcript of SFN was not quantifiable ($<7x10^{-4}$ n0/n0 HPRT) in MCF-7 cells, but it was determined in 18 of the 25 normal tissue samples in low amount (median: $8x10^{-5}$, Figure 29). Considering the different levels of determination, similar or even higher transcript levels in MCF-7 cells could not be excluded. However, on protein basis down-regulated SFN in human breast cancer cells compared to normal cells has been published previously (Vercoutter-Edouart et al., 2001).

Marker for proliferation and ESR1 activation

MCF-7 cells cultured with normal culture media are proliferating, thus higher levels of the proliferation markers were expected. The transcript level of MKI67, a well-known marker for proliferating cells (Chapter 2.1) was similar in all samples of normal tissue and MCF-7 cells (Figure 29). Hence, only levels of transcripts negatively modulating cell cycle were lower in proliferating cells (Figure 29). Since the normal culture media includes steroids, stimulating ESR1, higher levels of markers for ESR1 activation were expected. The transcripts of CCND1 and TFF1, two markers for ESR1 activation, were determined in a wide range in normal mammary gland tissue samples (Figure 29). While the transcript levels of TFF1 were higher in MCF-7 cells cultured with normal culture media (9.1 n0/n0 HPRT) than in cells cultured steroid-free (1.1 n0/n0 HPRT), the transcript levels of the second marker for ESR1 activation, CCND1, was slightly higher cells cultured with steroid-free media (1.6 n0/n0 HPRT) than in cells cultured with normal culture media (0.7 n0/n0 HPRT). The third marker for ESR1 activation, PGR, was determined in similar levels in all samples (Figure 29), indicating different sensitivities of the markers, since all three marker contain ERE or ERE-like sequences (Chapter 2.1).

Transcripts involved in E1/E2 metabolism

Qualitatively more isoenzymes of the E1/E2 metabolism were determined in human mammary gland tissue than in MCF-7 cells (Figure 29), which was independent of the culture conditions of the cells. The transcripts of HSD17B1, HSD17B2, CYP19A1, GSTP1, SULT1E1,

SULT1A1, UGT1A3/4, UGT1A8 and UGT2B7 were not determined in MCF-7 cells, but the higher limit of determination (20-fold higher than in normal tissue) must be taken into account. Five of these transcripts have been previously described as not expressed in MCF-7 cells: HSD17B1, CYP19A1 and SULT1E1 (Smuc and Rizner, 2009); UGT1A3/4 and UGT1A8 (Lehmann and Wagner, 2008). Contradictorily, mRNA of UGT1A3/4 has been detected in a previous study (Lehmann and Wagner, 2008), but was not detected in this study.

Both HSDs, HSD17B1 and B2, were determined in normal mammary gland tissue (median: 0.002 and 0.391 n0/n0 HPRT, respectively, Figure 29). The transcript levels of HSD17B2, catalyzing the conversion of E2 to E1 (Chapter 2.2.1), were 200-fold higher than the transcript levels of HSD17B1, catalyzing the conversion of E1 to E2 (Chapter 2.2.1), indicating higher E1 than E2 levels in normal tissues, which was in accordance to the determined E1 and E2 levels (determined by working group Lehmann, Chapter 5.2.14) in 22 of the 25 samples. In contrast, the transcript levels of HSD17B1 were below $7x10^{-4}$ n0/n0 HPRT (not determined) and of HSD17B2 were below $2x10^{-5}$ n0/n0 HPRT (not detected) in MCF-7 cells. Hence, both transcript levels were lower in MCF-7 compared to mammary gland tissue.

The transcript levels of CYP19A1, involved in local formation of E2 in the mammary gland (Chapter 2.2.1), were low in mammary gland tissue samples (median: 0.004 n0/n0 HPRT, Figure 29) and not quantifiable in MCF-7 cells ($<7x10^{-4}$ n0/n0 HPRT), thus lower in cancer cells than in normal tissue.

Hence, the formation of E1/E2 seemed to play a minor role in MCF-7 cells, since the three transcripts involved in formation (HSDs and CYP19A, Chapter 2.2.1) were either not detected or not quantifiable in cultured MCF-7 cells, while the three transcripts were determined in normal mammary gland tissue. Since all three transcripts have been described to be expressed in epithelial cells in normal tissue (Chapter 2.2.1) and the MCF-7 cell line was established from a breast adenocarcinoma, the loss of these transcripts was maybe due to dedifferentiation process of the tumor cells.

The transcript of GSTP1, involved in detoxification of reactive quinones (Chapter 2.2.2), was not detectable in MCF-7 cells ($<2x10^{-5}$ n0/n0 HPRT), while it was determined in normal mammary gland tissue, which is accordance to previously published data on the GSTP1 protein in MCF-7 cells (Terrier et al., 1990) and normal tissue (Howie et al., 1990). The transcript level of GSTP1 was the highest of all transcripts analyzed in this work in mammary gland tissues (median: 25.6 n0/n0 HPRT, Figure 29), which was in accordance with Oguztuzun et al. 2011, protein levels of GSTP1 were higher than GSTT1 levels in normal mammary gland tissues.

The transcript of SULT1A3/4 was the most abundant one of the SULT isoenzymes in mammary gland tissue (median: 0.1 n0/n0 HPRT, Figure 29) and the lowest expressed

transcript of the SULT isoenzymes was SULT1E1 (median: 0.002 n0/n0 HPRT, Figure 29). The transcript levels of SULT1A3/4 in normal tissue were similar to the levels in MCF-7 cells cultured with steroid-free media (0.2 n0/n0 HPRT, Figure 29), while the levels of MCF-7 cells cultured with normal culture media was 10 times lower (0.009 n0/n0 HPRT, Figure 29). The transcripts of SULT1A1 and SULT1E1 were not determined ($<7x10^{-4}$ n0/n0 HPRT) in MCF-7 cells (Figure 29), thus both levels were lower in MCF-7 cells, than in normal tissue.

All transcripts of the UGT isoenzymes were determined in very low levels in mammary gland tissue (Figure 29), except for UGT2B7, the transcript level ranged from $1x10^{-6}$ to 0.1 n0/n0 HPRT (Figure 29). The transcript of UGT2B7 was not detectable ($<2x10^{-5}$ n0/n0 HPRT) in MCF-7 cells, considering the higher limit of detection in MCF-7 cells, similar or even higher transcript level of UGT2B7 could not be excluded in MCF-7 cells compared to lowest transcript level in mammary gland tissue. Likewise, the transcripts of UGT1A3/4 and UGT1A8 were not quantifiable ($<7x10^{-4}$ n0/n0 HPRT) in MCF-7 cells, but with consideration of the limit of determination similar level in MCF-7 cells and normal tissue could not be excluded.

The transcript levels of STS in normal tissue (median: 0.8 n0/n0 HPRT) were comparable to the levels in MCF-7 cells cultured steroid-free (0.2 n0/n0 HPRT) and 100 times higher compared to the levels in MCF-7 cells cultured with normal culture media (0.009 n0/n0 HPRT, Figure 29). Hence, the transcript levels of STS in MCF-7 cells seemed to be influenced by the steroid level in culture media. In accordance, a decrease in STS activity of MCF-7 cells due to E2 treatment was reported previously (Tobacman et al., 2002).

The transcripts of COMT, CYP1B1, CYP1A1, GSTT1 and GCLC were similarly expressed in MCF-7 cells cultured steroid-free and normal tissue, while the transcript levels of the MCF-7 cells cultured with normal culture media were 5-10 times lower (Figure 29), indicating a steroid dependent transcription. A decrease of COMT and CYP1A1 mRNA levels and activities in MCF-7 cells have been observed previously for E2 (Lehmann et al., 2008; Wagner et al., 2008). Likewise, it has been described, that the mRNA level of CYP1B1 is regulated by E2, due to an ERE in the promoter region of the gene (Tsuchiya et al., 2004). Contradictorily, 12 h E2 treatment increased mRNA levels of CYP1B1 in MCF-7 cells (Tsuchiya et al., 2004), but no results were available on long time regulation of transcript levels of CYP1B1 by E2 in MCF-7 cells. It has been preciously described, that short term E2 treatment (48 h) promoted the process of Nrf2 translocation into the nucleus promoting antioxidant enzyme expression, e.g. GST and GCLC, in ESR1-positive myocardial cells (Yu et al., 2012), supporting the steroid dependent regulation of the transcripts of GSTT1 and GCLC observed in this study. The transcript level of NQO1, which can be also regulated by Nfr2 translocation (Chapter 2.3), had the highest values in mammary gland tissue (median: 1.9 n0/n0 HPRT). However, the transcript levels in MCF-7 cells cultured

steroid-free (0.4 n0/n0 HPRT) were lower than in cells cultured with normal culture media (0.8 n0/n0 HPRT), which was different to the transcript levels of GSTT1 and GCLC.

Nuclear receptors and paracrine signal transduction

As expected (Hanioka et al., 2012), MCF-7 were positive for ESR1, ESR2, AHR and ARNT (Figure 29). The transcript level of ESR1 was the higher in steroid-free cultured MCF-7 cells (3.3 n0/n0 HPRT) than in MCF-7 cells cultured with normal media (0.7 n0/n0 HPRT) and mammary gland tissue (median: 1.7 n0/n0 HPRT, Figure 29). Transcripts of NR1I2 and NR1I3 were not detectable ($<2x10^{-5}$ n0/n0 HPRT) in MCF-7 cells (Figure 29), in accordance to previous results (Hanioka et al., 2012). In contrast, all transcripts encoding receptors were determined in normal mammary gland tissues (Figure 29). The presence of transcripts of ESR1 and ESR2 in normal tissue were in accordance to previously published data (Chapters 2.1). No information on the presence of NR1I3 and ARNT in normal mammary gland tissue was available (Chapter 2.2.3). The transcript of AHR was detected in two epithelial cells lines derived form reduction mammaplasty (Chapter 2.2.3). NR1I2 has been described as not expressed in mammary gland tissue (Chapter 2.2.3). However, the transcript of NR1I2 was detected in each sample of mammary gland tissue in this study (Figure 29).

Levels of transcripts involved in paracrine signal transduction were higher in normal tissue than in MCF-7 cells (Figure 29). The transcript of WNT5a has been only detected in cultured normal epithelial cells, but no information on the presence in normal mammary gland tissue were available (Chapter 2.1), however, it has been detected in MCF-7 cells (Serra et al., 2011). The transcript level of WNT4 has been detected in normal mammary gland tissue, but not in MCF-7 cells (Huguet et al., 1994). Transcript levels of AREG and TGFB1 were expected in normal tissue (Chapter 2.1) and both have also been detected in MCF-7 cells (summarized in McBryan et al., 2008; Xie et al., 2011).

Conclusion

In conclusion, all of the analyzed transcripts, except for PGR and BAX, were differentially expressed in normal mammary gland tissue compared to MCF-7 cells. Additionally, the culture conditions of MCF-7 cells had influence on the transcripts involved in cell cycle and apoptosis, as well as on transcripts of enzymes involved in E1/E2 metabolism. In sum, MCF-7 cells possessed less mRNA of enzymes involved in E1/E2 metabolism and paracrine signal transduction concerning both number and amount of detectable transcripts. Since most of the transcripts were immunohistochemically detected in epithelial cells of normal mammary gland tissue (Chapter 2.2.2) and the MCF-7 cell line was established from a breast adenocarcinoma, the loss or reduction of transcripts or transcript levels were maybe due to dedifferentiation processes of the tumor cells. As expected, the number of detectable transcripts involved in cell cycle and apoptosis were similar in MCF-7 compared to normal tissue. Furthermore, increased proliferation, decreased apoptosis and differentiation

of the MCF-7 could be concluded, assuming comparability of transcript and protein levels. Additionally, transcripts involved in cell communication, such as receptors investigated in the study (except for ESR1) and proteins of paracrine signal transduction, were less abundant in MCF-7 cells than in normal tissue, presumably also due to dedifferentiation of the tumor cells. Since all transcripts of the E1/E2 metabolism were less abundant in MCF-7 cells than in human mammary gland, even if considering the different limit of detection, MCF-7 seemed to be rather inappropriate for the investigation of E1/E2 metabolism.

6.4 Preliminary establishment of transcript level quantification in epithelial cells

For investigation of the influence of factors influencing breast cancer risk on estrogen dependent action in human breast tissue it must be considered, that the mammary gland consists of many different cell types with different tasks an therefore different patterns of enzymes and signal proteins (Chapter 2.2). Hence, the effect of a factor may be different in each cell type. To investigate cell type specific expression, the objective was to establish a method for the quantification of transcript levels in specific cell types of the mammary gland.

Since the epithelial cells are the main sites of breast cancer development (Chapter 2.2), this work was focused on glandular epithelial and myoepithelial cells (GECs and MECs).

First, a method for isolating epithelial cells from the surrounding cells, such as stromal and adipose cells, was set up (Chapter 6.4.1). Then, the RNA quantity (Chapter 6.4.2) and the comparability of two MECs+GECs fractions (Chapter 6.4.3) were evaluated. In order to compare the gain of additional information by analyzing epithelial cells to information from homogenate analysis, the transcript levels of genes encoding enzymes in E1/E2 metabolism, apoptosis, cell cycle, paracrine signal transduction, proliferation and marker genes for proliferation (Chapter 6.2.1) were quantified and compared to the transcript levels in breast tissue homogenate (Chapter 6.4.4).

6.4.1 Laser capture microdissection

The aim was to investigate transcript level patterns in epithelial cells of the ductules since this is the main site of proliferation and differentiation (Chapter 2.2).

Since enzymatical methods cannot distinguish between epithelial cells of ductules or ducts and mRNA patterns are likely to be affected by this method of cell separation, the only method that enables specific isolation of epithelial cells of the ductules was laser capture microdissection. The method is based on preparation of sections, histological staining and microscopic examination of the tissue sections, which are placed on a special slide (Figure 30, Chapter 6.4.1). Tissue areas containing cells of interest (Figure 30) are directly cut with a laser beam and collected in a tube cap located under the slide.

Being the site of both proliferation (GECS) and signal transduction between epithelium and stroma (MECS), the aim was to isolate both, MECs+GECs.

RNA preserving treatment of the cells prior to RNA isolation and the number of the cells, respectively, are factors influencing the quality and quantity of the RNA. For a reliable spectrometrical quantification approx. 160 ng RNA was needed and for assessment of RNA integrity approx. 80 ng, since only approx. 0.18 ng RNA from 1000 isolated cells was expected (based on mean yield of RNA of homogenate, Table 61), RNA quantity and integrity

were assessed by quantification of the reference gene HPRT (Chapter 5.2.10).

Since the quality of the RNA could not be determined directly due to low amount, precautions were taken to avoid RNA degradation: To guarantee good quality RNA only mammary gland tissue aliquots were used which were flash frozen within 10-15 min after receiving the tissue sample from the operating room (Chapter 5.2.2). Then, the aliquots were stored at -80°C. The tissue aliquots were cut in sections (8 μm) with a cryo slicer at -20°C chamber temperature and -25°C tissue temperature, in order to keep the enzymatic activity of enzymes involved in RNA degradation as low as possible. Then, a quick staining of the sections was performed (Chapter 6.4.1). Since enzymes are active in an aqueous environment, the staining procedure excluded watering of the section and included a quick drying of the section after dipping the section into solutions with increasing ethanol content. After staining, the epithelial cells were immediately cut out.

For laser dissection of MECs+GECs the cutting lines were set outside of the MECs excluding the nuclei of stromal cells (Figure 31), to avoid contamination with RNA of stromal cells.

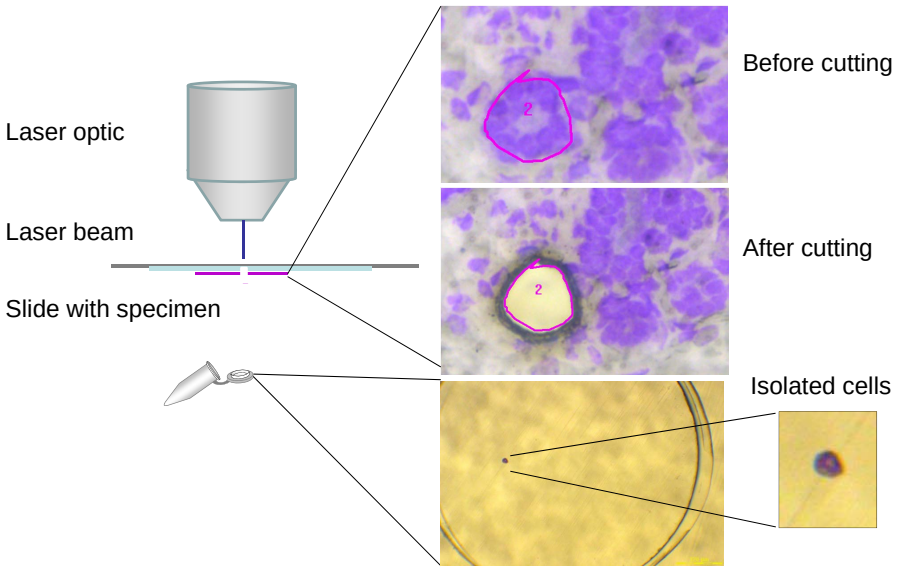


Figure 30: Scheme of laser capture microdissection of mammary gland. Kryo sliced (8 μm) mammary gland sections were stained with cresyl violet and cut with a Laser Microdissection system (Leica, Chapter subsubsec:LCM).

6.4.2 Quantity of RNA

Since processing time should be minimized (Chapter 6.4.1), determination of the exact number of the dissected cells was too time consuming.

Thus, for the estimation of the number of cells, number and area of laser dissected elements

were used. A dissected element (Figure 31) consisted of different number of cells, depending on size of cells and ductules.

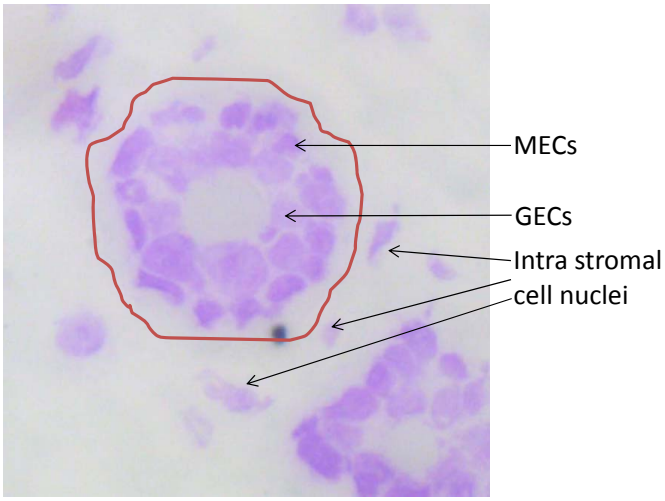


Figure 31: Scheme of setting the cutting lines for laser capture microdissection of GECs+MECs. For dissection of the MECs+GECs the cutting lines were set outside of the MECs excluding the nuclei of stromal cells.

In order to investigate, if the number and area of isolated elements were a suitable marker for the estimation of the cell number and consequently for quantity of RNA, MECs+GECs of eleven tissue sections (8 μm) of ten different tissue aliquots ($\widehat{=}$ ten different donors, selected due to quantity of of the tissue aliquots of good quality) were isolated. Then, RNA was isolated (Chapter 5.2.6), reversely transcribed and preamplified for 14 cycles (Chapter 6.2.4) and the n0 of HPRT was determined using TaqMan^{TM} gene expression assay (Chapter 5.2.10). Afterwards, the n0 was then plotted against the number of the laser dissected elements and area (Figure 32).

The transcript of HPRT was quantified in only ten of eleven samples, although the amount of elements (101) and area (198 mm^2) of the sample without detectable HPRT was comparable to the other samples, which had all comparable HPRT transcript levels. Surprisingly, neither the number of laser dissected elements (R=0.52, Pearson correlation coefficient), nor the laser dissected area (R=0.25) were well associated with n0 of HPRT (Figure 32). This could be due to the fact, that there was no "internal" standard for the quantification of the transcript of HPRT and thus no considering of e.g. loss of RNA during RNA isolation and different reverse transcription efficiencies of the samples. In addition, inter-individual differences could not be excluded, since cutting the same number of elements (130 and 131) with almost the same area (266 and 326 mm^2) of the same tissue aliquot, resulted in slightly higher n0 of HPRT of the cell fraction with the bigger area (64.3x10^3 and 103.7x10^3 copies, marked points in Figure 32). However, this suggest reproducible yields of RNA without

inter-individual influences.

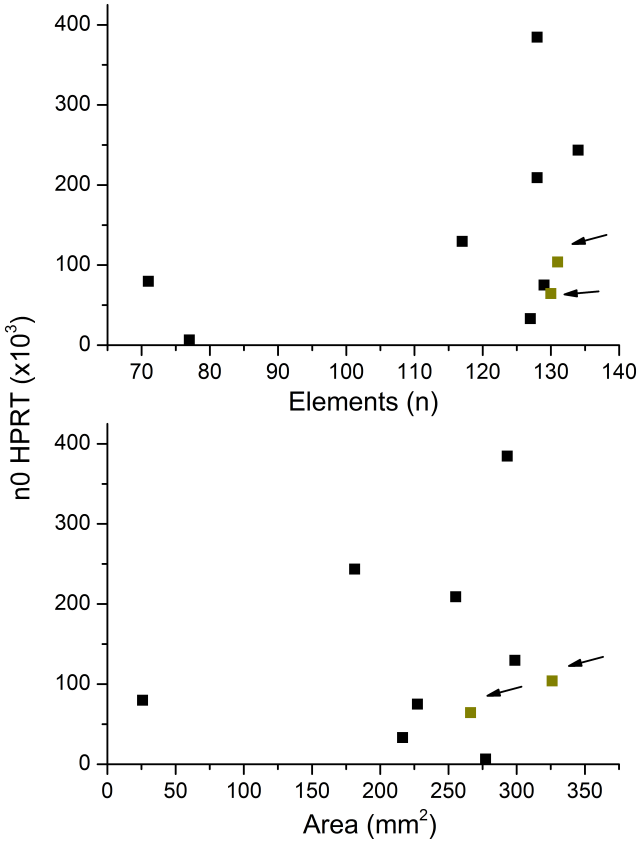


Figure 32: Number of laser dissected elements and area compared to n0 of HPRT of ten tissue sections (points) of nine different tissue aliquots ($\widehat{=}$ nine different donors). n0 of HPRT was determined in one (two tissue sections) or two dilutions (eight tissue sections) of the preamplified cDNA by means of TaqMan™ probe-based real time PCR. If two dilutions were used, the data are presented as mean. The arrows indicate n0 of HPRT of MECs+GECs of two tissue sections (8 μm) of one tissue aliquot.

6.4.3 Comparability of transcript levels in two MECs+GECs fractions

In order to investigate the technical reproducibility of mRNA levels determined in the study, cells were isolated using two serial tissue sections (#29I and #29II) from the same tissue aliquot. The transcript levels of both fractions were quantified using TLDA-PolymorphismII (Chapter 5.2.10) and standardized to reference gene HPRT. The n0 of HPRT transcripts were comparable in both fractions (n0=2.1x10^5 and 2.5x10^5).

25 of the 47 transcripts were quantifiable in MECs+GECs isolated from sample #29 (Figure 33). The variations in transcript levels ranged form 1% (range/2 of TP53) to 95% (range/2 of CDKN1A), most transcript levels varied between 20% and 50% (1-20%: 3, 20-

40%: 6, 40-50%: 6, 50-70%: 5, >70%: 4, Table A9.5). Only the transcript of BMF was quantifiifed in #29I, but not detected in #29II (Figure 33).

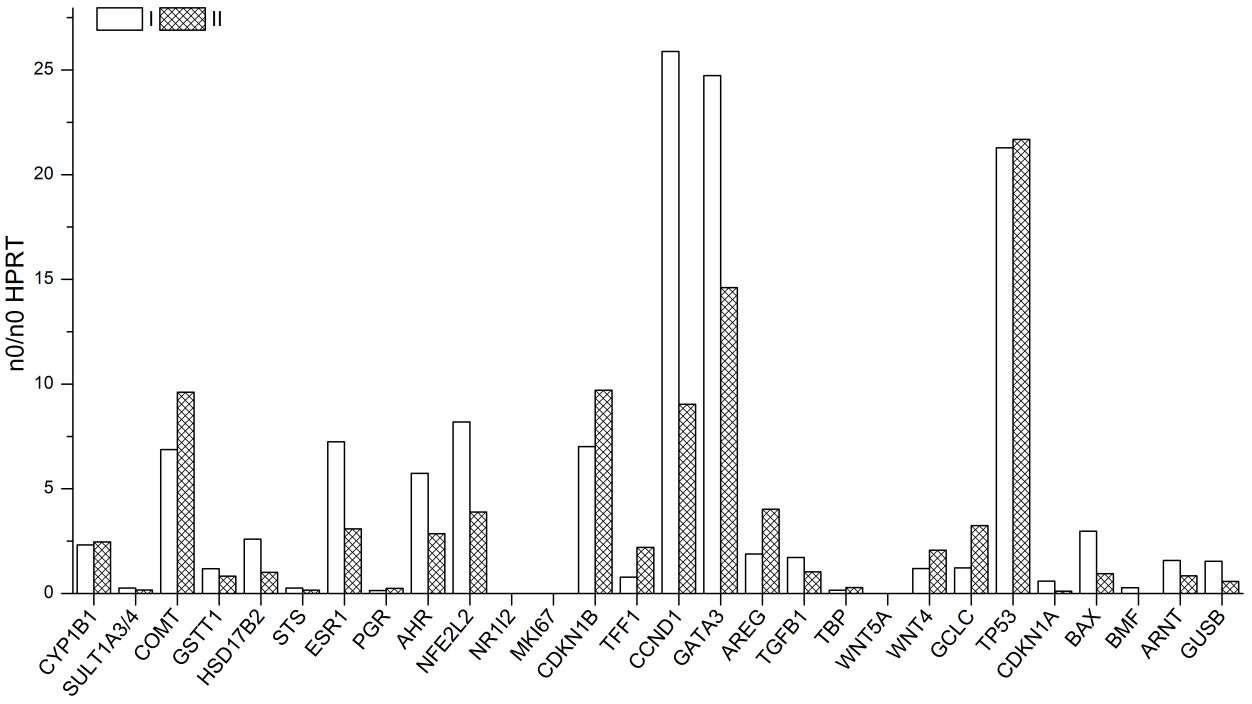


Figure 33: Transcript levels (n0/n0 HPRT) in isolated MECs+GECS of two tissue sections (I and II) of sample #29. Transcript levels were quantified by means of real time PCR (TLDA, Chapter 5.2.10, Table 6). cDNA of MECs+GECS were pre-preamplified (14+10 cycles, Chapter 5.2.10).

This suggested that the variations were not only dependent on the cell type but also on localization of the cells within the tissue. CDKN1A (highest variation) an important negative modulator of cell cycle progression (Chapter 2.1), would be expected in proliferation-active cells, which are 20% of the glandular epithelial cells (Chapter 2.1). Accordingly, the protein of CDKN1A was only detected in a few GECs, but in different distribution among the ductules (Chapter 2.1), causing maybe large variation in transcript levels in this work.

In contrast, the transcript of TP53 was quantified in all samples with low variation in high amounts. However, no positive signals of the TP53 protein neither in stroma, nor in epithelial cells were detected (Chapter 2.1). Maybe, the protein level in normal mammary gland tissue was under the limit of detection of immunohistology or the transcript level does not reflect the protein level, which is plausible since TP53 is regulated at the post-translational level by protein-protein interactions and covalent modifications (summarized in MacLaine and Hupp, 2011).

Since BMF is a mediator of involution (Chapter 2.1), the transcript would be theoretically expected only in apoptotic cells, which are presumably not equally distributed over the tissue.

However, no immunohistological staining of BMF in normal mammary gland tissue has been published until now (Chapter 2.1).

6.4.4 Transcript levels in epithelial cells and homogenate

MECs+GECs of nine tissue sections of eight different tissue aliquots were isolated using laser capture mircodissection (Chapter 6.4.1). Subsequently, total RNA was isolated and reversely transcribed. The cDNA was pre-preamplified (14+10 cycles, Chapter 6.2.4) and the transcript levels of genes encoding enzymes in E1/E2 metabolism, apoptosis, cell cycle, paracrine signal transduction, proliferation and marker genes for proliferation (Chapter 2.2) were determined with TLDA-PolymorphismII. To evaluate the putative gain of information by analyzing isolated MECs+GECs compared to analyzing homogenate, transcript levels were then compared to respective transcript levels of homogenate qualitatively and quantitatively.

Transcripts which are located in stromal and epithelial cells (HSD17B2 and STS, Chapter 2.2) as well as the reference genes were expected in similar levels in MECs+GECs and homogenate. Transcripts which are known to be exclusively expressed in MECs or GECs (GATA3, TFF1, ESR1, PGR, TGFB1, CDKN1A, CDKN1B, CCND1, CYB1B1, COMT, Chapters 2.1 and 2.2) based on immmunohistological staining or *in situ* hybridization, were expected in higher levels in MECs+GECs than in homogenate, because of lacking dilution with other cells not expressing these transcripts.

For quantitative comparison, the ratio of the transcript levels (n0/n0 HPRT) of MECs+GECs divided by the transcript level (n0/n0 HPRT) of homogenate were calculated. Due to the single determination and unknown variation, transcripts were only considered as higher or lower in MECs+GECs, if the ratio was $\geq$2.0 or $\leq$0.5, respectively. A ratio $>$0.5 and $<$2.0 was considered as equal.

For most transcripts (AREG, BAX, HSD17B2, PGR, SULT1A3/4, TFF1, ARNT, CCND1, COMT, CYP1B1, NFE2L2, CDKN1A) all three ratio variants (higher, lower, equal) were observed among the seven samples. For six transcripts (AHR, GSTT1, GUSB, STS, TBP, WNT5A) equal and lower levels, for six transcripts (CDKN1B, ESR1, GATA3, TP53, GCLC, WNT4) equal and higher levels and for the transcript of TGFB1 lower and higher levels were observed. The transcript level of BMF were lower and equal in all MECs+GECs samples compared to homogenate (Figure 34). No transcript was determined with higher level in all MECs+GECs samples (Figure 34).

As expected, the reference genes had similar transcript levels in MECs+GECs and homogenate, with the exception of sample #26, possibly indicating bias in any step of the quantification process.

Since at least one sample was completely different to the expectations, none of the lev-

els of target transcripts with known localization in mammary gland tissue corresponded to published immunohistological data (Chapters 2.1 and 2.2). However, some transcripts were partially matching the expectations of immunohistology: STS was expected at same levels in MECs+GECs and homogenate, but the transcript levels were equal or lower in MECs+GECs compared to homogenate. The transcript levels of CDKN1B, ESR1 and GATA3 were expected higher in MECs+GECs, but were equal or higher in the seven samples.

Since the ratios of the reference transcripts were in expected range for all samples, except for sample #26, an accurate determination of the target transcript levels was assumed. Excluding sample #26 from this evaluation changed only results of CCND1 and PGR: Both transcript levels of sample #26 had lower values in MECs+GECs compared to homogenate, whereas in the remaining samples both transcript levels were higher or equal in MECs+GECs (Figure 34), which was at least partially in accordance with the expected transcript levels for CCND1 and PGR (higher transcript levels in MECs+GECs).

One reason for the limited number of consensuses between the transcript levels and the immunohistological data might be the limit of detection of immunohistological staining, e.g. TP53 was not immunohistologically detected in normal mammary gland tissue (Chapter 2.1), but the transcript was quantified in all samples in this work, suggesting a higher limit of detection of immunohistology or a discrepancy of mRNA and protein level. Supporting this, it has been described that the human TP53 has 13 different splice variants in addition to the full length TP53 protein (Bourdon et al., 2005).

For the *in situ* hybridization of mRNA, another issue might be taken into account: The *in situ* hybridization of mRNA of CYP1B1 was carried out in five samples of normal tissue which were not further characterized concerning e.g. age, BMI, lobule type or smoking habit (Muskhelishvili et al., 2001). Likewise, the mRNAs of HSD17B2 and STS were detected by *in situ* hybridization in 25 samples of normal mammary gland tissue, but the samples were not further characterized or evaluated concerning person to person differences (Cavalieri et al., 2006). Furthermore, immunohistological staining of COMT of 19 normal mammary gland tissue samples were carried out, but person to person differences were not further addressed (Weisz et al., 2000). Hence, an influence of factors (personal or environmental) on the expression pattern of the transcripts is also possible.

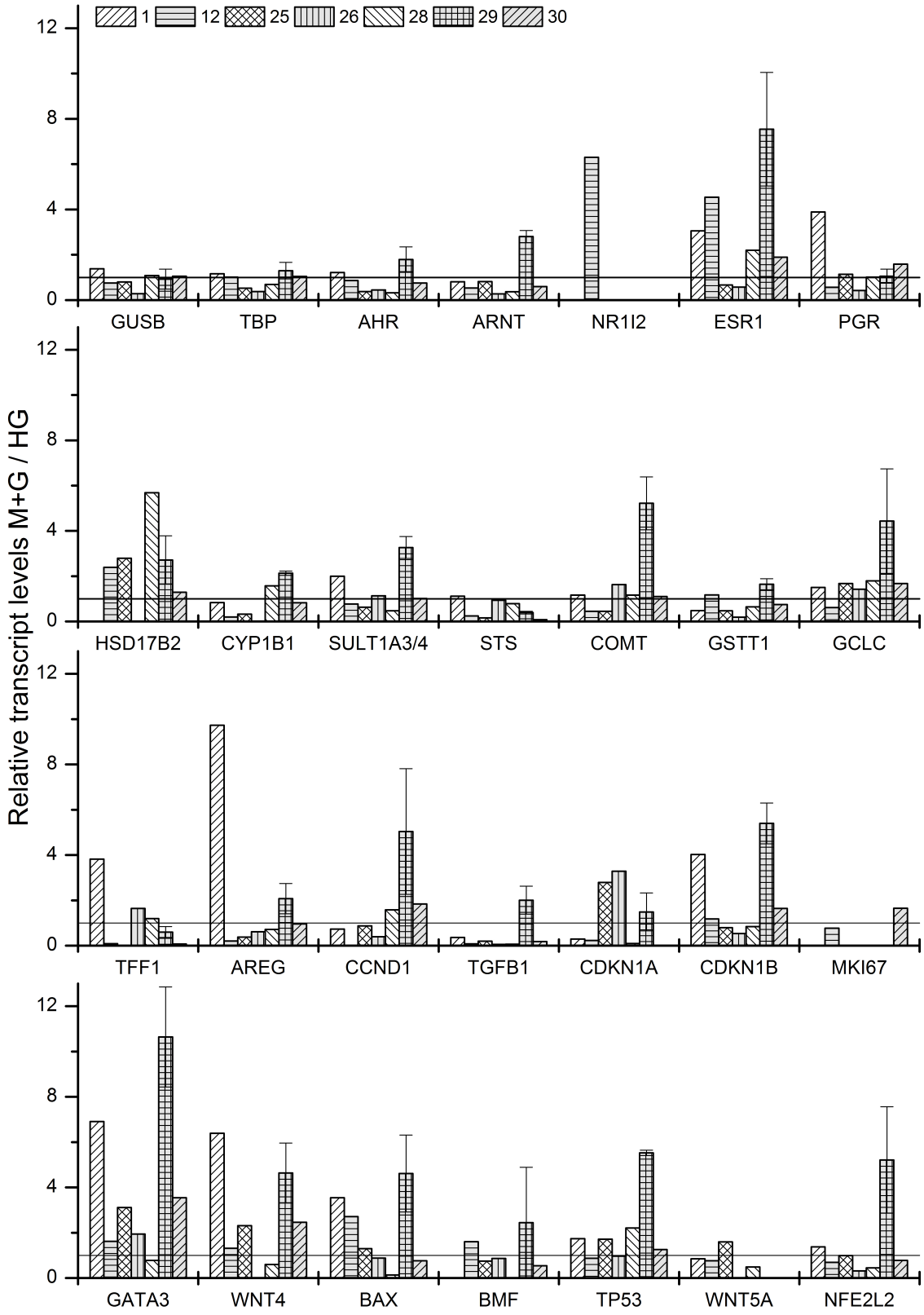


Figure 34: Transcript levels (n0/n0 HPRT) of MECs+GECS (M+G) relative to the respective transcript levels of homogenate (HG). Transcript levels were quantified by means of real time PCR (TLDA, Chapter 5.2.10, Table 6). cDNA of MECs+GECS were pre-preamplified (14+10 cycles, Chapter 5.2.10). The horizontal line indicates equal transcript levels in MECs+GECs and homogenate.

6.4.5 Comparison of transcript level pattern of MECs+GECs and homogenate

For quantitative comparison of the whole transcript level patterns observed in MECs+GECs to transcript level patterns observed in homogenate, a cluster analysis of transcript levels in homogenate and MECs+GECs of the same seven samples was performed (Chapter 5.2.16). Both for homogenate and MECs+GECs only transcripts were considered, which were detected in MECs+GECs in six of seven samples (Chapter 6.4.4), thus including ESR1, PGR, TGFB1, COMT and CYP1B1 (present only in epithelial cells), HSD17B2 and STS (present in epithelium and stroma), GSTT1, SULT1A3/4, AHR, AREG, ARNT, BAX, CCND1, CDKN1A, CDKN1B, GATA3, GCLC, GUSB, NFE2L2, TBP, TFF1, TP53 and WNT4 (unknown presence in stromal cells).

The cluster analysis revealed different cluster structures as well as different number of main clusters for homogenate and MECs+GECs.

Cluster homogenate

In the dendrogram of homogenate three main clusters were observed. However, cluster 3 was only based on sample #25 (Figure 35) and therefore this cluster was not taken into consideration for discussion. Cluster 1 comprised samples #12, #29, #30 and cluster 2 from samples #1, #26, #28.

The transcript levels of nuclear receptors, ESR1 and PGR, AHR and ARNT, and transcripts involved in ductal side branching, TGFB1 and GATA3, and transcripts involved in negative regulation of cell cycle, CDKN1A and CDKN1B, as well as the proliferation marker CCND1 were lower in cluster 1 than in cluster 2 of homogenate (Figures 35 and A74). In contrast, the markers of ESR1 activation in ESR1-positive GECs, AREG and TFF1, were similar in both clusters. Moreover, transcript levels of the signaling protein WNT4, were higher in cluster 1 than in cluster 2. Thus, homogenate clusters 1 and 2 did not differentiate between low and high ESR1 activation in ESR1 positive GECs, but between low and high proliferation/differentiation of ESR1 negative GECs.

Cluster 1 included all samples with a lobule type 2/3 and cluster 2 and 3 all samples with lobule type 1 or 1_{parous} (Figure 35). Russo and Russo (2004) described higher number of ESR1 and PGR positive (immunohistologically stained) epithelial cells in lobule type 1 compared to lobule type 2/3, which was also observed in these clusters (Figure 35).

Cluster MECs+GECs

Two main clusters were observed in the dendrogram of MECs+GECs (Figure 35). Cluster 1 was built from samples #12, #25, #26, #28, #30 and cluster 2 from #1, #29I and #29II. As expected, samples #29I and #29II were in the same cluster which indicates similar transcript level pattern (Figure 35).

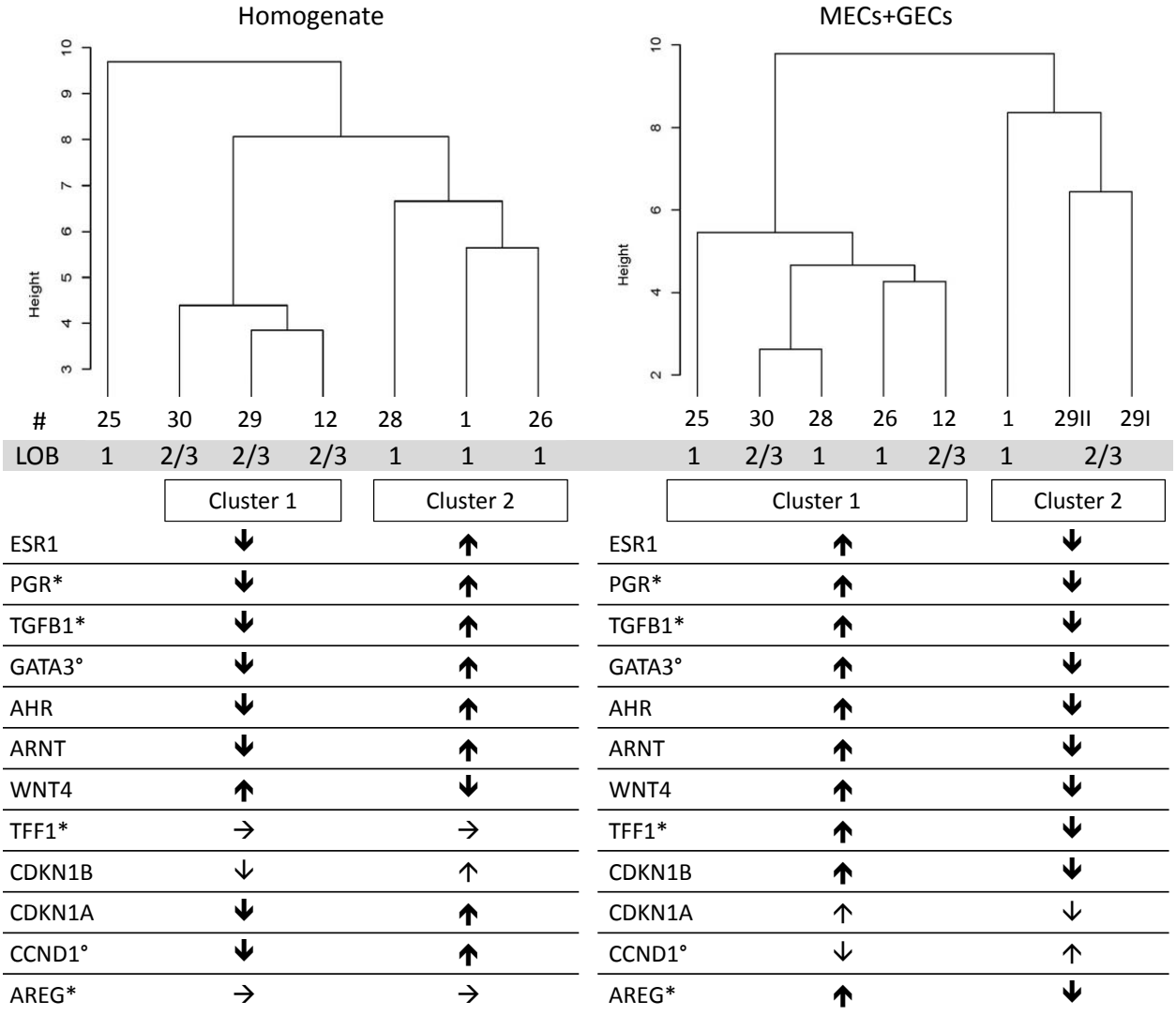


Figure 35: Cluster dendrograms of the smallest euclidean distance of transcript levels of homogenate and MECs+GECs of the samples. For cluster analysis only transcripts which were expressed in MECs+GECs of all samples were considered (Chapter 6.4.4). Additionally, due to huge differences in transcript levels, the transcript levels were standardized with "scale()"-function in R. Below sample numbers (#) the lobule type (LOB) is presented. The lower half of the figure presents the transcripts most influencing the clusters. The direction of the arrows indicate the transcript level in each cluster: up, high levels; down, low levels; right, equal level in both clusters. The thicker the arrows, the higher the transcript level difference between the clusters. *, transcripts containing ERE. °, transcripts ESR1-activation-mediated.

Cluster 1 of MECs+GECs was characterized by high levels of transcripts in ESR1 positive GECs (PGR, AREG, TFF1), of differentiation (GATA3, TGFB1), of negative cell cycle progression (CDKN1B, CDKN1A), of nuclear receptors (AHR, ARNT) and of the signaling protein WNT4 were high in cluster 1 and low in cluster 2 (Figures 35 and A75). Furthermore, transcript levels of the marker of indirectly induced proliferation of ESR1 negative GECs (CCND1) were elevated in cluster 1 and decreased in cluster 2, indicating tissues characterized by E2 activation but reduced proliferation and increased differentiation.

Thus, the samples were classified based on most of the markers for ESR1 activation in clusters of MECs+GECs and a decrease in proliferation and increase in differentiation. Yet, the two main clusters did not reflect the results of the histological lobule type determination (Chapter 6.1.2).

Different cluster composition in homogenate-based and MECs+GECs-based cluster analysis indicated different transcript level patterns of the signaling proteins WNT4 and AREG, and the proliferation marker CCND1 because of different cell compositions (Chapter 2.1). In the case indicating (i) occurrence of these transcripts in cell types besides MECs+GECs and (ii) differential regulation of their transcription ormRNA stability in MECs+GECs. Interestingly, lobule types were only reflected in clusters of homogenate, indicating contribution of expression of the analyzed transcripts in other cell types than MECs+GECs to lobule morphology.

These results were based on only seven samples due to discrepancies in transcript levels of reference genes in some dissected epithelial cell samples. To ensure more reliable applicability of the method to a larger number of samples, more optimization concerning RNA stability during isolation process is needed. Taken together, the mRNA levels of enzymes involved in E2 metabolism could not be analyzed in MECs+GECs due to low expression of the transcripts. In contrast, analysis of the same transcripts in homogenate and isolated MECs+GECs yielded comparable results for assessment of E2 activation but differed in the interpretation of proliferation. Furthermore, mammary gland histology could only be described with data derived from homogenate.

6.5 Influence of transcript levels and polymorphisms on genotoxic stress

The extent of genotoxic stress within the mammary gland tissue samples was assessed by DNA adduct fluxes calculated by bioinformatical network analysis (Chapter 6.5.1). Furthermore, the suitability of the Random Mutation Capture assay to determine frequencies of spontaneous mutations was investigated (Chapter 6.5.2).

6.5.1 Metabolic network

In order to interpret the individual differences in transcript levels of genes involved in E1/E2 metabolism and genotypes of polymorphisms (COMT Val108/158Met, NQO1 Pro187Ser, GSTT1 CNP) on E2 mediated genotoxicity (Chapter 2.2.2), the aim was to estimate the amount of DNA adduct formation.

Since no information on DNA adducts was available, a theoretical approach was chosen to calculate the DNA adduct formation by complementing the input data (transcript levels). This can be achieved by constructing a metabolic network (Chapter 2.3.1). The metabolic network was established with support of Prof. Dandekar and Dr. Cecil (Department of Bioinformatics, University of Würzburg).

For establishment of the metabolic network model, all reactions included in the E1/E2 metabolism and the network settings were defined first (Chapter 6.5.1.1). Then, the model was checked for plausibility to identify incorrectly annotated equations within the network (Chapter 6.5.1.2). Then, the model was calculated without and with real transcript levels quantified in this work, to investigate the influence of polymorphisms (Chapter 6.5.1.3) and individual mRNA levels (Chapter 6.5.1.4), respectively. In the last step, to evaluate the influence of polymorphisms considering individual mRNA level pattern, the network model was calculated using mRNA level patterns and genotypes of 25 samples (Chapter 6.5.1.5).

6.5.1.1 Reactions, settings and input data

Usually, all required pathways for the construction of the network are found in public data bases such as KEGG (Alexander Cecil, Department of Bioinformatics, University of Würzburg, personal communication, Chapter 2.3.1). However, the pathways of E2 and E1 metabolism (Chapter 2.2.2), were not completely annotated in KEGG data base: Only the reactions of the conversion of E2 to E1 and vice versa (catalyzed by HSD17B2 and HSD17B1, respectively), the conversion the E1- and E2-sulfate to E1 and E2 (catalyzed by STS), the hydroxylation of E1/E2 to 2HO-E1/E2 (catalyzed by CYP1B1) and the subsequent methylation of the catechols (catalyzed by COMT) were annotated. The transfer reaction

of a sulfo-group to E1/E2 and glucuronidation of E1/E2 were annotated but without the involved isoenzymes. In sum, only 14 of the 142 reactions needed to construct the network model were found in KEGG database. The remaining reactions were evaluated individually through a literature search, the result has been summarized in Chapter 2.2.2.

Additionally, to model co-factor production, energy metabolism (pathways of TCA cycle, oxidative phosphorylation, pentose phosphate pathway and glycolysis, Cecil et al., 2011) was added to the E1/E2 metabolism network. To simulate a normal human energy metabolism as flux constraints RNA sequencing data from Rowley et al. (2011) were used. The appropriate transcript numbers in the supplementary data of Rowley et al. (2011) for the KEGG database reaction numbers of the energy metabolism network were chosen. The reads per kilobase of exon per million mapped reads (RPKM) in the supplementary data of Rowley et al. (2011) were then used as flux constraints (input data). Three assumptions, however, had to be made:

1. If there was more than one transcript for a KEGG data base number, the sum of the RPKMs was used.

2. For 2 enzymes (alcohol dehydrogenase and dihydrolipoyl transsuccinylase) in the energy pathway no data was found in the supplementary data. The flux constraints for these were set to "1".

3. The flux constraints of the complexes of the oxidative phosphorylation were set to "1", because there were no data for the complexes in Rowley et al. (2011).

The next step was to define the settings for metabolites and cofactors for the calculation of the fluxes. There were two possibilities given by YANAsquare: internal (limited availability, restricted by the network) or external (unlimited availability, independent of network fluxes, Chapter 5.2.19). All metabolites connected to further metabolizing reactions were set as internal (e.g. catecholes and quinones, Table 36). Since E1 and E2 were the starting points and the subsequent network reactions should not be limited by availability of E1 and E2, they were set as external.

As cofactors NADPH (CYP and NQO1 reactions), SAM (COMT reaction), GSH (GSH conjugation), UDPGA (UGT reactions) and PAPS (SULT reactions) were considered. UDPGA and NADPH were set as external, since they are abundant in a normal cell (Eisenbrand et al., 2005). GSH was set as internal, since the mRNA of the GCLC, the key enzyme of the GSH biosynthesis (Richman and Meister, 1975) and regulated through cellular stress response (Chapter 2.2.2) was quantified (Chapter 6.3). SAM and PAPS were set as internal, since they are of limited availability in the human cell (Eisenbrand et al., 2005). The settings for cofactors and metabolites are summarized in Table 36 and the settings for energy metabolism were published in Cecil et al. (2011).

6.5.1.2 Plausibility of the model

In order to check plausibility of the model, flux constraints of all reactions were set to "1". The calculation of the model, however, failed due to unachievable constraints. Thus, minimum one reaction needed to be set to "0", to have more possibilities for a constraint conform solution of the calculation (Chapter 2.3.1). Since UGT1A3/4 was not expressed in any of 25 analyzed samples (Figure 28), it was set to "0".

Since the E1 and E2 metabolisms contained the same reactions with only slight differences in isoenzymes of a few reactions (e.g. transfer of a sulfo group to E1 and E2, Figure 6), no differences in the flux value of the same enzyme were expected for E1 and E2. Only negligible differences in the flux of three reactions were observed (Figure 36):

1. The flux of the transfer reaction of a sulfo group to E1 and E2 were 1.0 and 1.3, respectively.

2. The flux of the glucuronidation reaction of 2-MeO-E1 and 2-MeO-E2 were 0.1 and 0.2, respectively.

The most active reactions were those catalyzed by CYP1A1 and CYP1B1. The oxidation reactions of E1 and E2 to the catechols catalyzed by CYP1B1 were twice as high as the reactions catalyzed by CYP1A1. This is plausible since CYP1B1 catalyzed both the reactions to 2-HO-E2/E1 and 4-HO-E2/E1, whereas CYP1A1 only catalyzed the reactions to 2-HO-E2/E1 (Chapter 2.2.2). This also explained that reactions of 2-HO-E2/E1 to the quinones and DNA adducts were more active compared to 4-HO-E2/E1, since there were two reactions synthesizing 2-HO-E2/E1. The second most active reactions were those catalyzed by COMT. All these reactions were producing metabolites which were further metabolized, since the more reactions a metabolite is involved in, the more active the synthesis reaction (Chapter 2.3.1). All reactions producing final metabolites (in context of this network, e.g. GSTs, SULTs and UGTs) had lower activities. The sum of reactions catalyzed by UGT isoenzymes had higher activities than SULT or GST catalyzed ones. This is most likely due to the limited cofactor production for SULT and GST (reactions restricted to 1 vs. unlimited availability).

Hence, all reactions were annotated correctly and the metabolic network model was considered as plausible.

6.5.1.3 Impact of polymorphisms on respective enzyme fluxes

In order to investigate the impact of the polymorphisms associated with an increased breast cancer risk (NQO1 Pro187Ser, GSTT1 CNP) and possible interactions of polymorphisms which were not associated with an increased risk (COMT Val108/158Met, Chapter 2.4.1) on

DNA adduct formation, polymorphisms were included in the network calculations (Figure 36) by multiplying the the flux constraints (1) with the relative activity of the respective enzymes: For COMT Val108/158Met the activity for Met/Met enzyme is only 50% of the Val/Val enzyme (Chapter 2.4.1), thus the factor was 0.5.

For the heterozygous genotype no activity has been described yet, therefore an intermediate activity was assumed (75% Val/Met of Val/Val, factor 0.75, Chapter 2.4.1). The NQO1 Ser/Ser enzyme has an activity of 2% of the Pro/Pro enzyme (factor 0.02, Chapter 2.4.1, Traver et al., 1997). Since for the heterozygous genotype no activity has been described yet, again an intermediate activity was assumed (49% Pro/Ser of Pro/Pro, factor 0.49). Since the CNP of GSTT1 directly influenced the amount of mRNA (Figure 17) the factor for the homozygous type of GSTT1 was set to "1", for hemizygous to "0.5" and nullizygous to "0" for calculating the model with test data sets.

Calculating the model with consideration of the COMT Met/Met ("dCOMT") and the Val/Met ("dCOMT_05") genotype resulted in 75% and 87%, respectively, of the COMT Val/Val flux (Table 44). Likewise, the NQO1 Ser/Ser ("dNQO1") and NQO1 Pro/Ser ("dNQO1_05") genotype resulted in 19% and 57%, respectively, of the NQO1 Pro/Pro flux. The nullizygous genotype of GSTT1 ("dGSTT1") resulted in a flux of 0 for the GSTT1 reactions, while simultaneously the fluxes of the second and third GST isoenzymes, GSTP1 and GSTM1, reactions (Table 44) were increased (133%) which lead to a decrease of the total GST flux to 89%.

Hence, the modifications of flux constraints due to genotypes of COMT Val108/158Met and NQO1 Pro187Ser were reflected in the calculated fluxes of the respective enzymes, but not in full range. In contrast, both the nullizygous and the hemizygous genotype of GSTT1 resulted in a flux of GSTT1 of 0.

For estimation of the influence of the polymorphisms on DNA adduct formation the calculated fluxes of DNA adduct formation were compared. Only NQO1 Ser/Ser and Pro/Ser genotypes resulted in an increase in DNA adduct forming flux of 48% or 24%, respectively (Table 44). In contrast, the COMT Met/Met or Val/Met and the nullizygous/hemizygous genotype of GSTT1 had no influence on DNA adduct forming fluxes (Table 44). One explanation could be the different effects of the genotypes on the whole network. While the COMT Met/Met genotype resulted only in slight changes in the calculated fluxes of COMT and the GSTT1 -/- and +/- genotypes were almost compensated by GSTP1 and GSTM1 fluxes. In contrast, NQO1 Ser/Ser and Pro/Ser genotypes resulted in substantial differences in NQO1 flux which were not compensated by network effects.

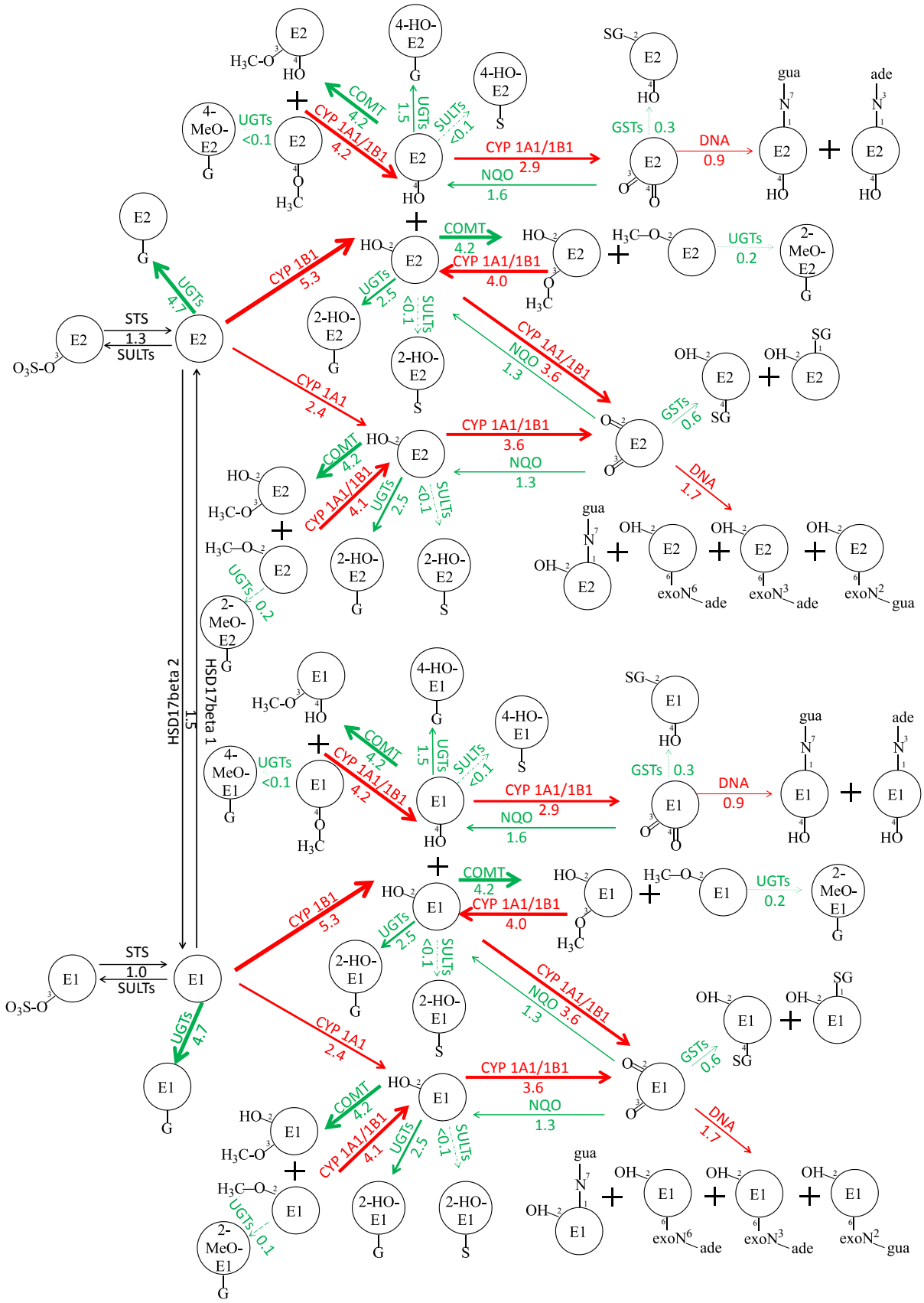


Figure 36: Metabolic network calculation of plausibility check. Flux constraints for all reactions were set to "1", except UGT1A3/4 which was set to "0". Single fluxes (arrows, line thickness proportional to value of flux) were presented as percentage (number under each arrow) of total flux sum of all reactions involved in E1/E2 metabolism. The reactions of the energy metabolism are not shown in this figure.

Table 44: Calculated fluxes of COMT, NQO1, GSTP1, GSTT1, GSTM1, sum of GSTs and formation of DNA adducts for the different genotype scenarios with decreased detoxification (PM variants). For calculation of the flux values without considering polymorphism genotypes, all flux constraints (FC) were set to "1", except UGT1A3/4 was set to "0" (d). The flux values of the genotype with decreased detoxification are presented as percentage of the respective flux values of d (%F_d). The FC of COMT, NQO1 and GSTT1 of each genotype scenario with decreased detoxification were set as indicated and the remaining reactions were set as in d. Each FC correspond to a genotype: 1.00, homozygous COMT or NQO1 or GSTT1; 0.75, COMT Val/Met; 0.50, COMT Met/Met or GSTT1 +/-; 0.49, NQO1 Pro/Ser; 0.02, NQO1 Ser/Ser; 0.00, GSTT1 -/-. n.a., not applicable. dMIN_DEAKT, minimal activity of COMT, GSTT1 and NQO1.

Enzyme/adducts	**PM variants**													
	dNQO1		dNQO1_05		dCOMT		dCOMT_05		dGSTT1		dGSTT1_05		dMIN_DEAKT	
	FC	%F_d	FC	%F_d	FC	%F_d	FC	%F_d	FC	%F_d	FC	%F_d	FC	%F_d
COMT	1.00	100.0	1.00	100.0	0.50	74.5	0.75	87.0	1.00	100.0	1.00	100.0	0.50	74.4
NQO1	0.02	19.0	0.49	56.9	1.00	100.0	1.00	100.0	1.00	100.0	1.00	100.0	0.02	19.0
GSTP1	1.00	114.3	1.00	100.0	1.00	100.0	1.00	100.0	1.00	133.3	1.00	133.3	1.00	133.3
GSTT1	1.00	100.0	1.00	100.0	1.00	100.0	1.00	100.0	0.00	0.0	0.50	0.0	0.00	0.0
GSTM1	1.00	100.0	1.00	100.0	1.00	100.0	1.00	100.0	1.00	133.3	1.00	133.3	1.00	133.3
$\sum$ GST	n.a.	114.3	n.a.	100.0	n.a.	100.0	n.a.	100.0	n.a.	88.9	n.a.	88.9	n.a.	88.9
DNA adducts	1.00	148.0	1.00	124.0	1.00	100.0	1.00	100.0	1.00	100.0	1.00	100.0	1.00	148.0

Furthermore, the combination of genotypes (COMT Met/Met, NQO1 Ser/Ser and GSTT1 -/-) with the theoretically lowest detoxification of reactive E1/E2 metabolites was calculated. The resulting fluxes of the respective enzymes were equal to those with only one genotype of one enzyme changed ("dMIN_DEAKT", Table 44).

In conclusion, of the three tested polymorphisms only NQO1 Pro187Ser had an impact on the DNA adduct formation flux in the metabolic network model. The effect of the nullizygous genotype of GSTT1 was almost compensated by a higher flux of GSTP1 and GSTM1, which resulted in slightly lower total GST (89%) flux and in no change in DNA adduct formation fluxes. The COMT Val108/158Met polymorphism affected the fluxes of the COMT reactions only slightly (decrease to 75%) and had no further influence on the fluxes of the network.

6.5.1.4 Impact of mRNA levels on DNA adduct formation fluxes

In order to investigate the impact of the transcript levels of genes encoding enzymes involved in E1/E2 metabolism on the DNA adduct formation fluxes in the metabolic network, transcript levels (Chapter 6.3) of all enzymes involved in E1/E2 metabolism of all samples were used as flux constraints for calculation of the network fluxes for each sample.

The flux of DNA adduct formation ranged from <0.0004 (Sample #0) to 0.037 (Sample #8, Figure 37) indicating a high influence of individual mRNA levels.

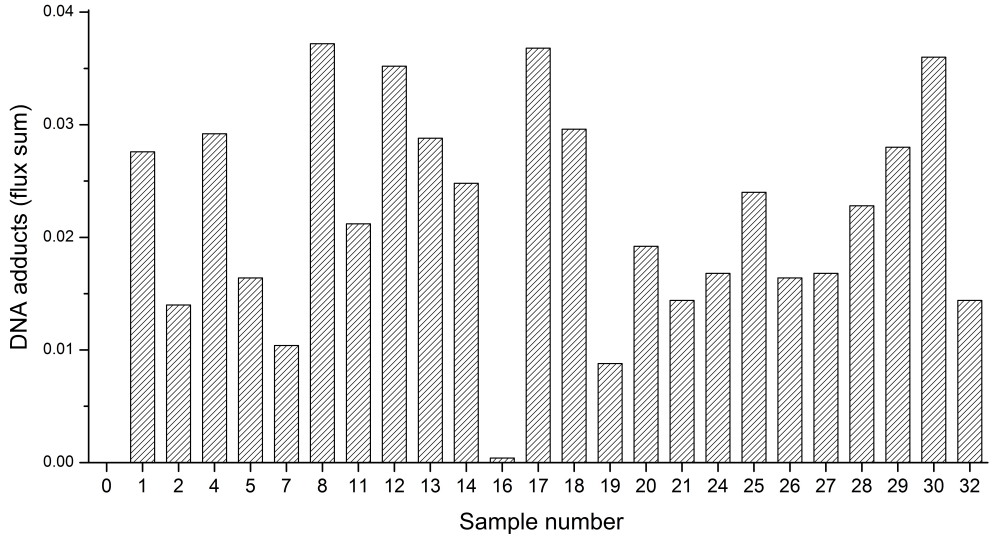
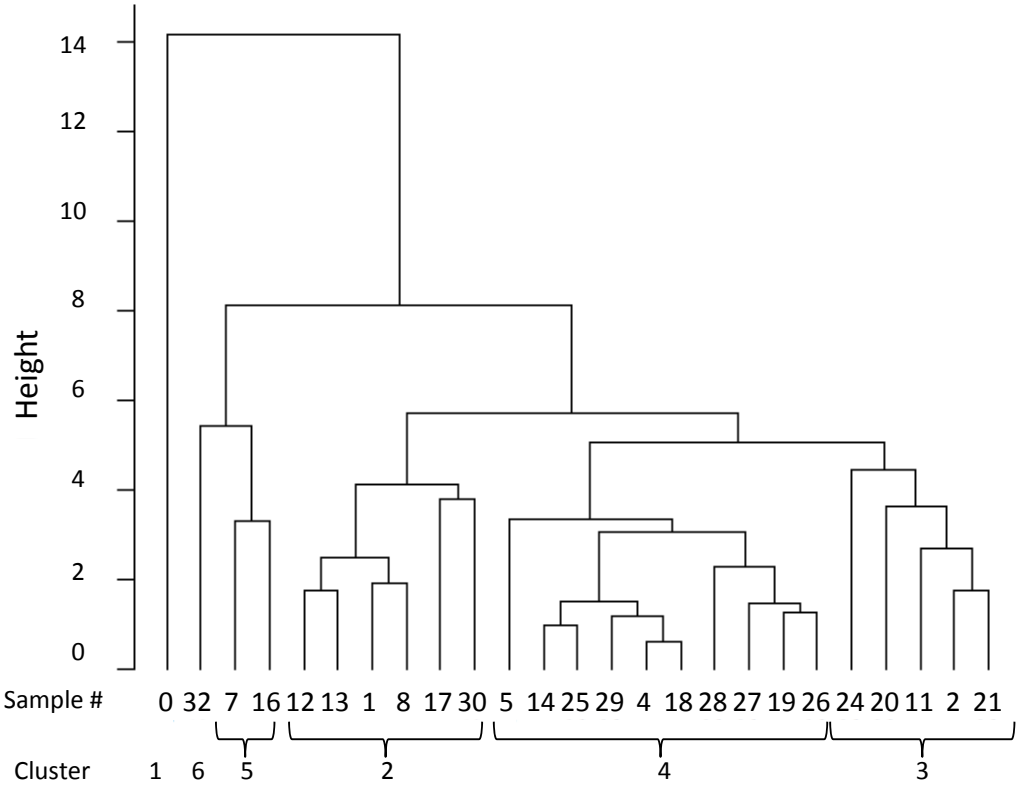


Figure 37: Sum of E2 and E1 DNA adduct formation fluxes calculated with transcript levels of 25 mammary gland tissue samples presented as absolute flux values. Flux constraints were set to individual transcript levels of all enzymes involved in E1/E2 metabolism quantified by means of real time PCR (Chapter 6.3) for each sample. The flux constraints of DNA adduct formation were set to "5".

In order to investigate which fluxes of enzymes involved in E1/E2 metabolism are decisive

for these huge differences in DNA adduct forming fluxes a cluster analysis (Chapter 5.2.16) of all samples with calculated fluxes of enzymes directly metabolizing E1 and E2 with flux values >0 (CYP1A1, CYP1B1, SULT1A1, SULT1A2, SULT1E1, SULT2A1, COMT, GSTP1, NQO1, HSD17B1, HSD17B2, STS, GCLC) and E1/E2 DNA adduct forming fluxes was performed. As expected, sample #0 and sample #8 were in different clusters indicating differences in their pattern of fluxes (Figure 38). Furthermore, sample #0 formed its own cluster while sample #8 formed a cluster together with samples #1, #12, #13, #17, #30 (Figure 38) which also had high DNA adduct fluxes (0.028, 0.035, 0.028, 0.036, 0.036, respectively). There was one more cluster consisting of sample #7 and #16 which also had low DNA adduct forming fluxes. This cluster showed the the same characteristics even though the absolute differences were smaller (Figure 38).

Comparing the two clusters with low DNA adduct formation fluxes to clusters with high ones the former had remarkably higher flux values in GCLC and GSTP1 (Figure 38). Thus, the flux values of both enzymes seemed to play a major role for DNA adduct formation fluxes in this network. Both enzymes are involved in detoxification of reactive E1/E2 metabolites: GSTP1 catalyzes the conjugation of very reactive quinones with GSH while GCLC is the key enzyme of the GSH synthesis (Chapter 2.2.2).



DNA adduct	↓	↓	↓	↑	→	↓
GCLC	↑	→	↑	↓	→	↑
GSTP1	↑	→	↑	↓	→	↑
NQO1	↑	→	→	↓	↓	→
COMT	↑	↑	↑	→	↓	→
SULT1A1	↑	↑	↑	→	↓	↓
SULT1A2	↑	→	↑	→	↓	↓
SULT2A1	↑	↑	↑	→	↓	↓
SULT1E1	↑	↑	↑	→	↓	→
STS	↑	↑	↑	→	↓	→
CYP1A1	↑	↑	↑	↓	↓	↑
CYP1B1	↑	→	↑	→	↓	↑
HSD17B1	↑	→	↑	→	→	→
HSD17B2	↓	↑	↓	→	↓	→

Figure 38: Cluster dendrogram of the smallest euclidean distance of calculated fluxes of DNA adduct formation and transcripts. For cluster analysis fluxes of DNA adduct formation and transcripts indicated were considered. Additionally, due to huge differences in fluxes, they were standardized with "scale()"-function in R. The lower half of the figure presents fluxes influencing the cluster. The direction of the arrows indicate fluxes in each cluster: up, high flux; down, low flux; right, equal flux. The thicker the arrows, the higher the fluxes. Important fluxes are highlighted in black.

6.5.1.5 Impact of mRNA levels and polymorphisms on DNA adduct formation fluxes

Since the DNA adduct forming fluxes were dramatically influenced by the individual mRNA pattern of the E1/E2 metabolizing enzymes, the influence of the polymorphisms on DNA adduct forming flux considering real mRNA levels of 25 samples was investigated. Genotypes of all samples for COMT Val108/158Met and NQO1 Pro187Ser (Table 52) were included in the flux constraints by multiplying transcript levels with the respective factors as described in Chapter 6.5.1.4. For the GSTT1 CNP no factor was included, since GSTT1 -/+ resulted in significantly decreased transcript level of GSTT1 compared to GSTT1 +/+ and the transcript level of the GSTT1 -/- genotype was not detectable (Chapter 6.1.3.3).

Since the high-activity genotypes of COMT and NQO1 required no adjustment of the flux constraints, no changes in DNA adduct forming fluxes were expected for the five samples with this genotype. In addition, the DNA adduct formation fluxes of other 14 samples did not change due to "genotype-adjustment" of flux constraints. These samples were characterized by medium- or low-activity genotype of COMT combined with high-activity genotype of NQO1, except for sample #12, the genotype of which was high-activity COMT combined with medium-activity NQO1 (Figure 39). An increase in DNA adduct formation of different quantity was observed in six samples (Figure 39). These were all samples with a genotype of medium-activity NQO1 combined with medium- or low-activity COMT. The highest increase was observed for sample #24 (82.6%) and the lowest increase for samples #13 (1.4%, Figure 39 and Table A67). Interestingly, the genotype of both samples was the same (COMT Met/Met and NQO1 Pro/Ser) only the mRNA levels of the E1/E2 metabolizing enzymes were different, indicating strong influence of individual transcript patterns. However, even with an increase to 182.6% the absolute DNA adduct forming flux of sample #24 was not as high as the one of sample #13.

In sum, only for samples with medium-activity genotype of NQO1 combined with medium- or low-activity genotype of COMT, increased DNA adduct formation fluxes were observed. However, there was only one sample with medium-activity genotype of NQO1 and high-activity genotype of COMT, where no increase in DNA adduct formation was observed. Hence, to draw conclusion more samples with this specific genotype are needed.

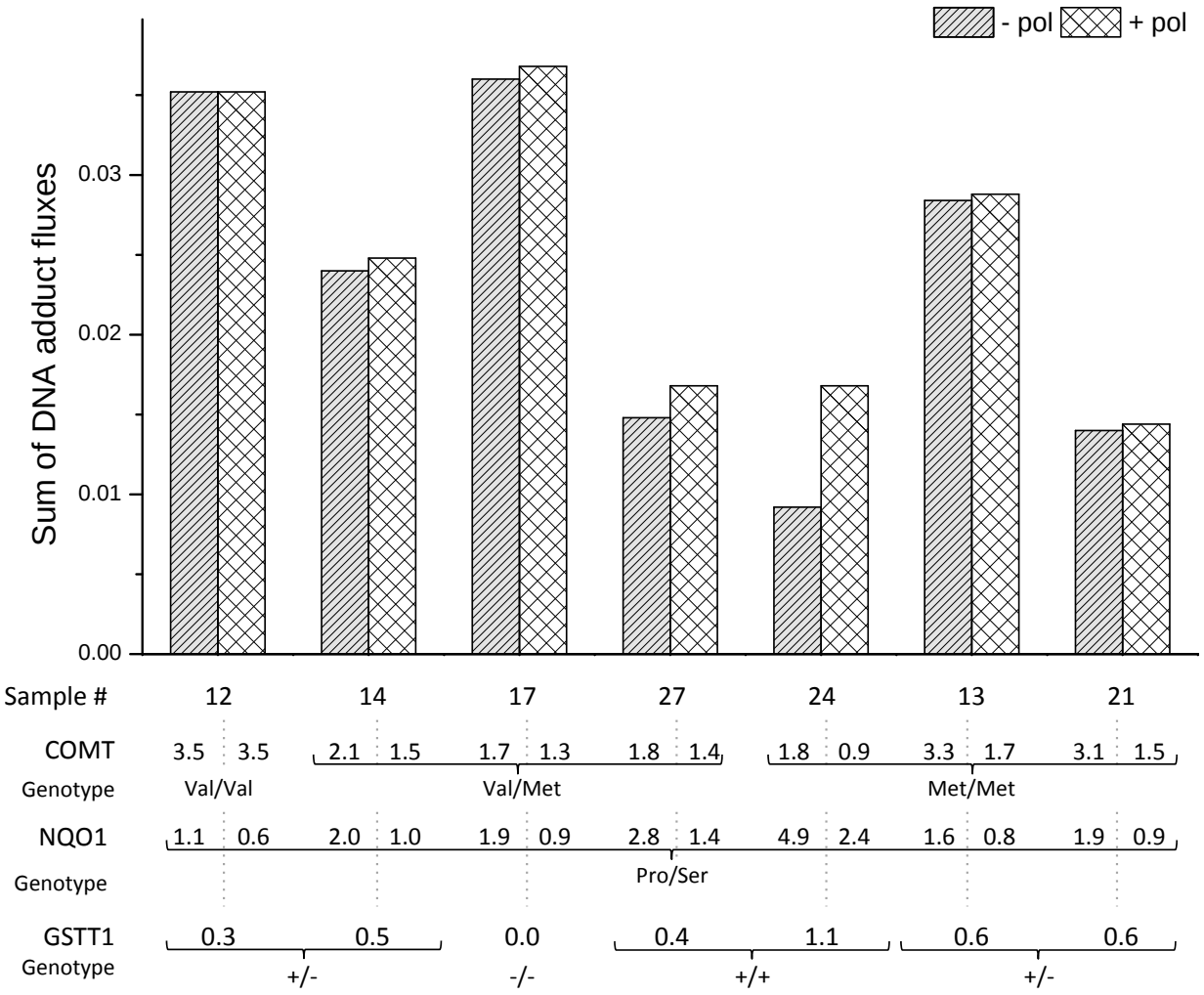


Figure 39: Sum of DNA adduct formation fluxes of E2 and E1 without and with consideration of genotypes. Presented are absolute flux values. The setting of the flux constraints for COMT, NQO1 and GSTT1 without (-pol) and with (+pol) consideration of the individual genotype are described below x-axis. Grey bars (-pol) and white bars (+pol) The transcript levels of all enzymes involved in E1/E2 metabolism quantified by means of real time PCR of all samples were set as flux constraints. The flux constraints of the DNA adduct formation reactions were set to "5".

6.5.1.6 Limitations and discussion of the metabolic network

The metabolic network presented a simplified model of the E1/E2 metabolism in the breast tissue. Despite some limitations of the network model (Chapter 6.5.1.6.1), changes in the balance of toxifying and detoxifying pathways due to transcript levels or genotype differences were detected (Chapter 6.5.1.6.2).

6.5.1.6.1 Limitations of the metabolic network

Because of the simplifications of the reactions, the current model of the mammary E1 and E2 metabolism has four limitations:

Availability of E1 and E2

E1 and E2 are unlimitedly available which does not reflect the real situation in the tissue (Chapter 2.2.2). This was set up, since the aim was to investigate the influence of changes in the metabolism on the DNA adduct levels, under the steady-state conditions, thus in a system without other perturbations or extracellular influences. Thus the quantitative levels of E1 and E2 were of no interest. Additionally, to construct a model of the whole network which calculates E1/E2 metabolites quantitatively several other factors need to be considered (described next), which were not possible to include in this kind of simplified network model.

Enzymatic parameters

Although the model incorporates all known isoenzymes (e.g. UGTs and SULTs), it does not take into account, if two or more enzymes/isoenzymes competing for one metabolite, which enzyme/isoenzymes is favored. This could only be realized in a more detailed network, where specific parameters of each isoenzyme are taken into account. For the prediction of the favored enzymes, a factor needs to be calculated using two enzyme parameters k_{cat} and K_M (Eisenthal et al., 2007). However, k_{cat} and K_M are not available for all enzymes/isoenzymes involved in E1/E2 metabolism and would need to be determined experimentally.

Inhibiting reactions

Possible inhibitions of enzyme activites by metabolites were not considered, e.g. the feedback inhibition of the methylestrogens on the CYP1A1 and 1B1 activity (Chapter 2.2.2). Like the first two limitations this could be considered in a more detailed network which includes the enzymatic parameters.

Transport reactions

The network only displays the metabolism in one cell. Transportation reactions, influx or efflux of metabolites or E1/E2 are not considered in the current model which may also influence the whole network. However, to model the transport reactions more information on the transporter (e.g. OATPs or OATs, Chapter 2.2.1) is needed (e.g. transcript levels) which were also currently not available and would need to be determined experimentally.

6.5.1.6.2 Discussion of the metabolic network

The NQO1 Pro187Ser SNP is associated with an increased breast cancer risk (about 27%, Chapter 2.4.1). Assuming increased calculated DNA adduct fluxes lead to more DNA adducts in the cells, then the mutations may also increase, which in turn may lead to an increase in the number of initiated cells and thus to a higher breast cancer risk. Hence in this discussion, the calculated DNA adduct forming fluxes are assumed as markers for breast cancer risk. The DNA adduct fluxes calculated without consideration of real mRNA pattern of the E1/E2 metabolizing enzymes for the NQO1 Pro187Ser SNP met the epidemiological results, since the DNA adduct forming flux was increased to 148% (NQO1 Ser/Ser, low active) compared to the medium active NQO1 Pro/Pro. However, when considering real quantitative mRNA patterns, only the combination of medium-activity genotype of NQO1 with the medium- or low-activity genotype of COMT increased DNA adduct formation fluxes. Yet, there was only one sample with medium active NQO1 Pro/Ser and high active COMT Val/Val. Until now there are no epidemiological studies published concerning a combination effect of these two polymorphisms on breast cancer risk. Other combinations of polymorphisms in E1/E2 metabolizing enzymes are known to be associated with an increased risk of breast, e.g. COMT Val108/158Met and CYP1B1 Leu432Val (Cerne et al., 2011).

The COMT Val108/158Met SNP is not associated with breast cancer risk (Chapter 2.4.1), which matches the results of this network model concerning the DNA adduct formation fluxes.

In a recent meta-analysis GSTT1 CNP was slightly associated with an increased breast cancer risk (about 11%, Chapter 2.4.1), but no increase in DNA adduct formation fluxes were observed with this network model, due to compensation effects of the other GST isoforms. The relevance of this network effect remains unclear, since no information on the affinity of the GST isoenzymes towards the E1/E2 metabolites has been published until now. Thus, it cannot be deduced if the GST isoformes can replace each other functionally.

Since not all effects met results of previous studies, it must be taken into consideration –among other limitations of this model– that the influence of the polymorphism on E1/E2 metabolism is only one possibility to increase the risk of breast cancer, since all of the three enzymes (COMT, NQO1 and GSTT1) are also involved in detoxification of other reactive metabolites (endogenous and exogenous) which may influence breast cancer risk. Despite this, the model is capable of showing the influence of polymorphisms alone or in combination on the whole E1/E2 metabolism.

6.5.2 Random Mutation Capture Assay

. DNA adducts do not always cause mutations (Figure 8). Thus, another approach was the determination of mutation frequencies in mammary gland tissue. Since histological normal tissue was to be analyzed, the spontaneous mutation frequency (SMF) was expected to be very rare ranging from $3x10^{-7}$ to $1x10^{-8}$ (Chapter 2.3.2). Hence, a method applicable to human mammary gland tissue and sensitive enough to determine low SMFs in normal tissue was needed.

Therefore, the RMCA a genotype selective method which detects mutants that render the mutational sequence non-cleavable by the *Taq*I restriction enzyme after accumulation of the target sequence (Chapter 2.3.2) was tested for its suitability to investigate the SMF of human normal mammary gland tissue in this study.

First, the method was performed as described previously with slight modification (Chapter 6.5.2.1) to determine the SMF of MCF-7 cells and normal mammary gland tissue (Chapter 6.5.2.2). Then, optimization of the PCR steps of the RMCA were performed for saving time and financial resources (Chapter 6.5.2.3).

6.5.2.1 RMCA of TP53

The RMCA is divided into eight steps (Chapter 2.3.2) which can be summarized into six practical steps:

1. Isolation and digestion of genomic DNA with restriction enzymes.

2. Hybridization of DNA with an uracil containing biotinylated DNA probe and enrichment of the hybridized target.

3. Digestion of target DNA with *Taq*I and digestion of the uracil containing probe with UDG.

4. Quantitative PCR to determine the total number of target genes.

5. PCR to determine mutant frequency.

6. Verification of mutant by resistance to *Taq*I digestion.

All steps were published for the intron 6 of human TP53 gene (Bielas and Loeb, 2005), but had not been applied to normal human breast tissue. Thus, all steps, especially the DNA isolation and the amount of tissue, had to be adapted.

Isolation and digestion of genomic DNA with restriction enzymes

Since the expected SMF was 1x10^{-8} per base pair (Chapter 2.3.2) and with every *Taq*I restriction site four bases are investigated, at least 2.5x10^{7} copies of the target sequence were necessary to detect at least one mutant copy. 2.5x10^{7} copies of the target correspond to approx. 100 μg genomic DNA (Equation (2), Chapter 2). DNA of approx. 2 g mammary gland tissue was isolated by an extended proteinase K digestion (approx. 48 h) prior to chloroform extraction (Chapter 5.2.7).

With this method 104±67 μg (Range: 19-378 μg) intact DNA per 1 g mammary gland tissue (n=34 tissue aliquots) were isolated and the quality determined by agarose gel electrophoresis (Figure 40, lane b). 104 μg corresponded to a DNA yield of 3x10^{7} copies/g tissue. Since one mutant copy is not enough to calculate a reliable SMF, the aim was to detect of more than one mutated copy. In approx. 350 μg (1x10^{-8} copies) four mutant copies were expected. Accordingly, the required amount of mammary gland tissue for the determination of the SMF was at least 4 g. Based on the 30 samples collected in this work, approx. 50-400 g of tissue per breast per reduction surgery were gained for this study and even with a very high amount of adipose tissue (Chapter 6.1), 4 g of mammary gland tissue with 2-20% attached adipose tissue were available of every sample, thus the modified DNA isolation step was applicable for the tasks of this study.

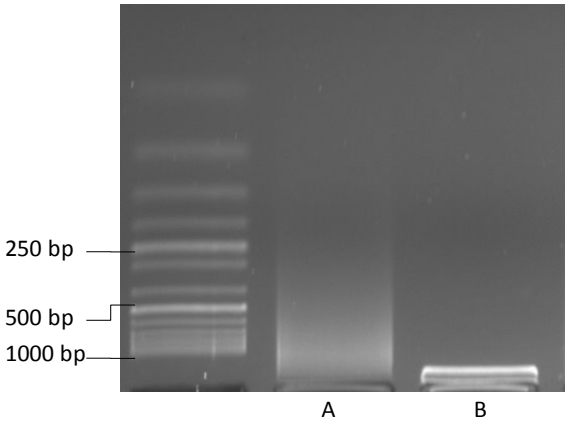


Figure 40: Electrophoretic separation of intact genomic DNA, which did not migrate in electric field (b) and digested DNA (restriction enzymes: *Eco*RI, *Eco*RV, *Pvu*II, *Bam*HI, *Rsa*I for 16 h at 37°C) (a) on an 1.5% agarose gel stained with Sybrgreen®. Left lane, DNA ladder.

Thus, approx. 350 μg (1x10^{-8} copies) DNA were digested with five different endonucleases (*Eco*RI, *Eco*RV, *Pvu*II, *Bam*HI, *Rsa*I) for 16 h at 37°C to eliminate steric problems in the following hybridization step. The digestion with the five restrictions enzymes resulted in fragmented DNA, as expected (Figure 40, lane a).

Hybridization of DNA and accumulation of the hybridized target

The target sequence in intron 6 of TP53 gene was captured by hybridization with a complementary, uracil containing, 5'-biotin-terminated DNA-probe, synthesized via PCR, followed by magnetic separation from the remaining DNA as described in Chapter 5.2.12) analogous to Bielas and Loeb (2005).

Digestion of target DNA with *Taq*I

The captured target sequences were digested five times with *Taq*I to assure that mutations in the uracil containing probe do not lead to false positives (Bielas and Loeb, 2005). Then, the uracil containing probe was digested with UDG, in order to avoid the probe serving as template in the subsequent PCR. Theses steps were carried out analogous to Bielas and Loeb (2005).

Quantitative PCR to determine the total copy number of target sequences

The method published by Bielas and Loeb (2005) used a Sybrgreen® based real time PCR method for quantification of the copy numbers of the captured target sequences. In this work copy numbers were quantified by means of competitive PCR, since this method was more unsusceptible to matrix interference of the PCR. For competitive PCR an internal standard, which could be amplified with the same primers as the target sequence, but yielded a PCR product of different length was synthesized by means of PCR (Chapter 5.2.12) and the number of copies was determined fluorimetrically using Sybrgreen® (Chapter 5.2.12.7). After competitive PCR, the standard product (180 bp) was electrophoretically separated from the copy number product (250 bp, Figure 41). With an increasing amount of internal standard compared with constant amount of the copy number product a calibration curve was determined and the copy number calculated (red arrows, Figure 41).

PCR to determine mutant frequency

The number of mutant copies is determined by splitting all copies to a number of single PCRs in a way that only one mutant copy can be expected to be present in a reaction. Thus, every reaction yielding a PCR product with the primers specifically amplifying mutant copies (Chapter 2.3.2, Figure 9), can be assumed to be caused by one mutant copy.

According to Poisson distribution one mutant copy per well is guaranteed, if ten PCRs contain a single mutant copy and 69 PCRs do not contain a mutant copy. Thus, copy numbers per well were adapted for each sample individually, according to the expected SMF. Usually, about $1x10^5$ copies per each reaction were analyzed in a series of 79 PCRs. If more than ten mutant copies were identified in 79 reactions, a second series of PCRs with a lower number of total copies per reaction was set up.

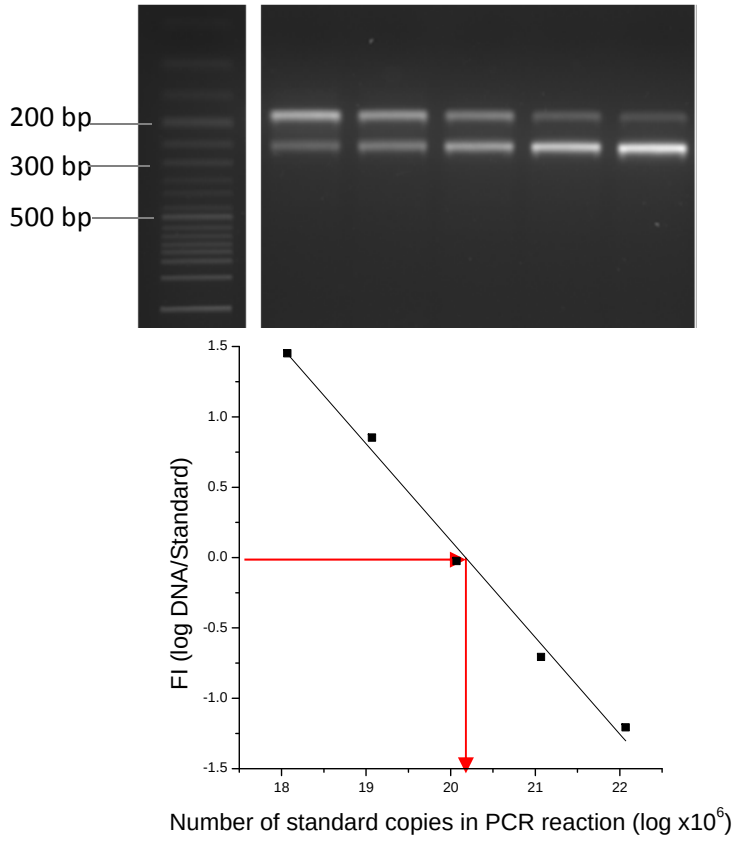


Figure 41: (A) Electrophoretical separation of the products of competitive PCR (internal standard, 180 bp and copy number product, 250 bp) on a 3% agarose gel stained with Sybrgreen®. (B) Graphical analysis (linear equation) of competitve PCR. Red arrows indicate the copy number of the sample.

Verification of mutants by resistance to *Taq*I digestion

Incomplete *Taq*I digest is known to lead to amplification of wild type copies (false positive mutant reactions). Therefore, amplification products of PCRs with the primer pair specifically amplifying mutant copies (Figure 9) were again digested with *Taq*I (Post-*Taq*I-digest, Figure 42B) resulting in cleavage of amplification products of wild type copies. Resistance or susceptibility of amplification products to *Taq*I digest was visualized by agarose gel electrophoresis (Figure 42B). Thus, false positive amplification products could be identified and excluded from SMF calculation.

The reproducibility of the mutant copy number quantification was assessed by performing three series of 71 reactions amplifying mutant copies out of one *Taq*I digested sample. The number of mutant copies detected in each run were 9, 8 and 9. The calculated SMF (Chapter 5.2.12) of each run ($4.24\text{x}10^{-7}$, $3.77\text{x}10^{-7}$ and $4.24\text{x}10^{-7}$, respectively) resulted in a relative SD of SMF ($4.08\pm0.22\text{x}10^{-7}$) of 5.4%. Hence, it was sufficient to detect mutants only once

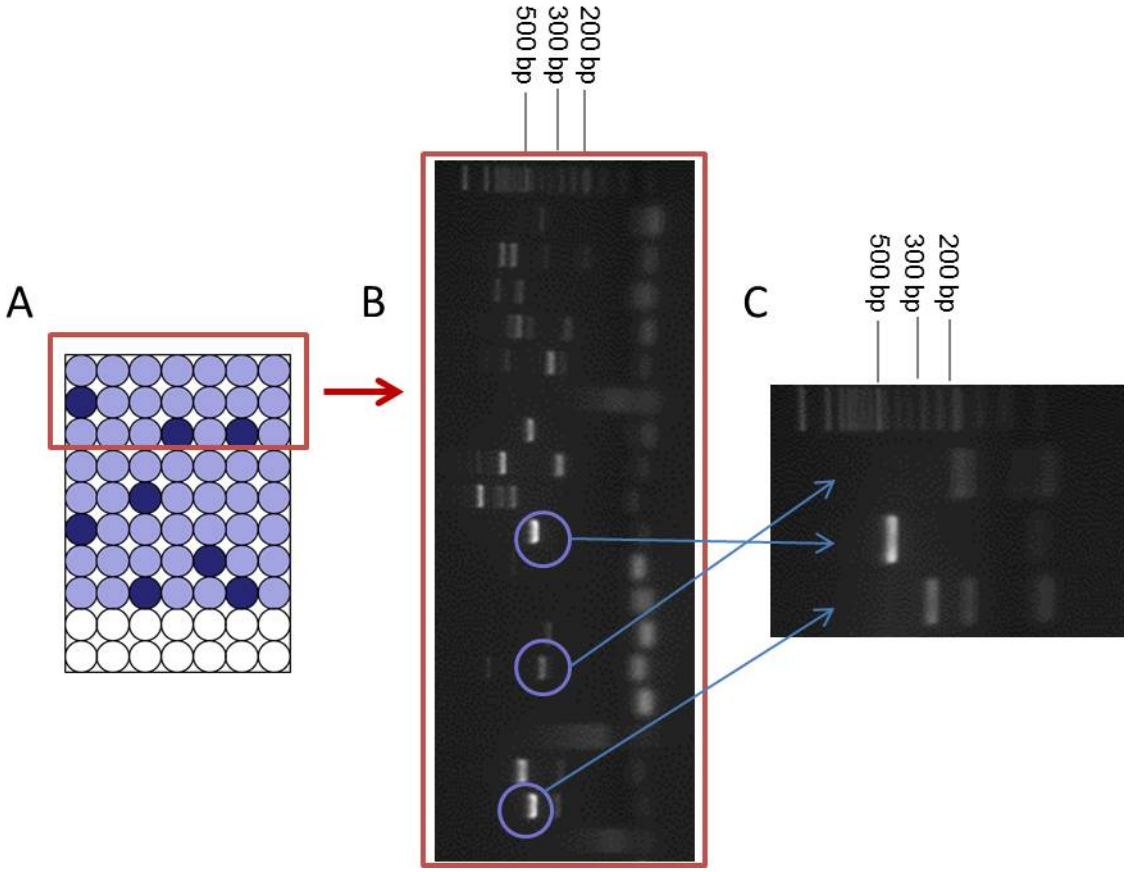


Figure 42: Scheme of mutant detection procedure. A: Scheme of 80 PCRs (including one no template-control) in a 96-well-plate to detect number of mutant copies. Light blue circles indicate wild-type target copies. Dark blue circles indicate mutant target copies. B: To detect mutants an electrophoretical separation of all PCR products on a 3% agarose gel stained with Sybrgreen® was performed. Positive mutants were post digested with *Taq*I. Blue circles, exemplary positive mutants (band at 418 bp). C: Electrophoretical separation of *Taq*I-post-digest on a 3% agarose gel stained with Sybrgreen®. The lower lane presents a false positive mutant (no band at 418 bp, but two shorter bands). The middle lane presents a mutant product (signal at 418 bp). The upper lane presents a secondary PCR product.

in one *Taq*I digest.

6.5.2.2 SMFs of cell lines and normal human mammary gland tissue

To check plausibility of the results, a comparison of the SMF of mammary gland tissue to the SMF of cultured cells were made, since no SMF of normal mammary gland tissue had been described. With the method described above, the SMF in intron 6 of TP53 was detected of two different breast cancer cell lines (MCF7 and MDA), of two different samples of normal human mammary gland tissue and of primary fibroblasts (AG0 cells). The SMFs of cultured cancer cell lines were expected to be higher and the SMF of cultured primary cells similar to the SMF of normal tissue (Prindle et al., 2010).

As expected, both cancer cell lines MCF7 and MDA had very high SMFs in intron 6 of the TP53 gene of 3.7x10^{-6} and 6.9x10^{-6}, respectively. The SMFs of human mammary gland tissues (3.6x10^{-7} and 7.2x10^{-8}) were 10 to 100-fold lower than the SMFs of the cancer cell lines. The SMF of the primary cell line (6.9x10^{-7}) was comparable to the SMFs of normal tissues, as expected. Additionally, the SMF determined in normal mammary gland tissue was comparable to the SMF in the intron of FEN1 gene in normal human lung tissue (3x10^{-7} Zheng et al., 2007). No further data on SMF in human intron sequences either in normal or cancer tissue or cultured cells has been published until now. In conclusion, the method was sensitive enough to detect SMF of normal human breast tissue, but for 30 samples, collected in this study, the methods is too time consuming and cost-intensive.

6.5.2.3 Optimization of the RMCA

In order to reduce time and financial resources for the detection of the SMFs in normal mammary gland tissues in this study, the PCR based quantification of numbers of total and mutant copies were optimized.

6.5.2.3.1 Quantification of copy numbers

The hitherto used competitive PCR method to determine the number of total copies, which needs gel electrophoretic separation of amplification products, was time consuming and resource-intensive. Since Sybrgreen® based real time PCR was assumed to be too unspecific for reliable quantification, a specific TaqMan® probe-based method with external calibration curve was to be established.

For this purpose, a primer pair amplifying a short product (134 bp) to ensure a high efficiency of the reaction and a matching TaqMan® probe was designed (Table 45).

Table 45: Primer established for TaqMan® probe based real time PCR detection of TP53 copy number. Primer were verified by size and sequence of the amplification product (Appendix).

Forward primer	5'-GCCCTCCAGGTGAGCAGTAG-3'
Reverse primer	5'-AGGCCCTTAGCCTCTGTAAGCT-3'
TaqMan®probe	5'-FAM-ACCTCCCTATAACCCCATGAGATGTGCAA-BHQ1-3'
Length product	134 bp

To establish the quantification method, a standard template to be used for the generation of the calibration curve was generated and quantified. Then, the linear range and statistical parameters of the calibration curve were determined. At last the reproducibility and accuracy of sample analysis was evaluated.

Quantification of the standard template

For external calibration, a standard template for the calibration curve is required. It was necessary, that the standard template contained the sequence of the primers and TaqMan® probe and was similar in length and bp sequence to the target sequence. The DNA probe, used for enrichment of the target sequence, met these requirements. Thus, the DNA probe was synthesized by means of PCR but without uracil. Then, the PCR product was quantified by fluorescence spectrometry (Chapter 5.2.12.7), aliquoted and stored at -20°C. The content of the standard was 28.0±2.2 ng/μl (1.9±0.2x10^{11} copies/μl) and sufficient for quantification of 300 samples.

Linearity and recovery rate

In order to identify the linear range of the calibration curve three independent serial dilutions of the standard from 1x10^2 to 2x10^7 copies were analyzed with the real time PCR based method. The linearity of the curve was tested with the software VALOO 2.5.

The range from 2x10^4 to 2x10^7 was linear and the variances were homogeneous. Additionally, no outliers were detected in this range.

To determine the recovery rate without influence of the matrix, a solution of DNA probe (without uracil) with a known concentration (2.5*10^6 copies per reaction) was diluted three times independently and quantified with the real time PCR based method.

The concentration of the solution was 2.47±0.3x10^6 copies which corresponded to a recovery rate of 98.8±8.0%. Thus, both the accuracy as well as the reproducibility of the method was ascertained.

Table 46: Summary of the method for copy number determination for enriched target sequence the TP53 gene, using real time PCR and external calibration curve. Data presented as mean±SD.

Range	2x10^4-2x10^7 copies/reaction
Recovery rate	98.8% ± 8.0%
Slope of regression line	-3.3
R^2	0.984-0.999 (n=8)

6.5.2.3.2 Mutant detection

Analogous to copy number detection, for saving time and financial resources the PCR based mutant detection was optimized. To save the first gel electrophoresis, the aim was to detect the mutants with Sybrgreen® based real time PCR, keeping hitherto primers and reagents.

First, an appropriate Sybrgreen® concentration not inhibiting PCR was identified. Since the first gel electrophoresis was no longer necessary for mutant detection, the volume could be reduced to 10 μl and the series of reactions could be carried out in a 384-well plate.

Sybrgreen® concentrations from 0.2 μM to 20 μM in reaction mix were tested and the melting curve of the amplification products were compared with their length determined by agarose gel electrophoresis.

No detectable amplification curves were observed in reactions with Sybrgreen® concentrations of 2-20 μM and of 0.02μM. Additionally, no PCR product was observed after gel electrophoretic analysis of the respective reaction mixtures containing 2-20 μM Sybrgreen®, while in the reaction containing 0.02 μM a band was detected (data not shown). Thus, Sybrgreen® inhibited the PCR reaction at concentrations above 2 μM and 0.02 μM Sybrgreen® was not concentrated enough to generate a detectable amplification curve. In contrast, 0.2 μM Sybrgreen® resulted in a detectable amplification and melting curve as well as a band detectable after agarose gel elctrophoresis.

6.5.2.3.3 SMF in MCF-7 cells

The optimized protocol was used to determine the SMF of another batch of MCF-7 cells. The result ($6.7x10^{-6}$) was 80% higher than the SMF determined with the published protocol in another MCF-7 cell batch ($3.7x10^{-6}$, Chapter 6.5.2.2), which was probaby due to different numbers of population doublings during cell culture.

6.5.2.3.4 Summary and Discussion

Using the optimized method, it was possible to determine the number of total copies of more than one sample in one PCR run without loosing accuracy and reproducibility. The optimized detection of the mutant copies saved time, since it saved one gel electrophoresis step, and financial resources, because of the omitted electrophoresis step the volume of the each reaction could be scaled down to 10 μl. However, the hybridization efficiency was often too low to collect a sufficient number of copies to contain at least 4 mutant copies (about $1x10^8$ copies). Furthermore, the sensitivity of the PCR reaction amplifying mutant copies was not constant enough to amplify a single mutant copy in every series of PCR reactions.

This caused time-consuming and expensive repetitions of these steps. Thus, for the amount of samples which are needed for statistical evaluation, time and funds required for the determination of the SMF using the RMCA were too unpredictable to further pursue this task.

6.6 Statistical evaluation of factors interacting with E1/E2 metabolism

Factors correlated with breast cancer risk could interact with E1/E2 metabolism (Chapter 2.4), thus influencing E1 and E2 tissue levels in the mammary gland which in turn could affect proliferation by ESR1 activation (Chapter 2.1), differentiation or apoptosis within the mammary gland. Hence, the aim was to identify correlations between transcript levels quantified in the work and E1 and E2 tissue levels as well as DNA adduct formation fluxes considering factors associated with breast cancer risk.

Therefore, the associations of these factors with transcript level patterns determined in homogenates of the 25 mammary gland samples were statistically investigated.

To identify correlations among transcripts of genes encoding enzymes E1/E2 metabolism, apoptosis, cell cycle, paracrine signal transduction, proliferation and marker genes for proliferation (Chapter 2.2.2) with similar regulation (e.g. by ESR1 activation), a correlation analysis was performed (Chapter 6.6.1).

Since the were human data with extreme values which may bias the correlation coefficient, the rank-based Spearman correlation coefficient was used. The correlation analysis was also used to identify associations between transcript levels and factors influencing breast cancer risk (Chapter 6.6.2). However, correlation analyses is only applicable for continuous variables such as age and BMI. For statistical analysis of categorical variables, such as smoking habit, a multiple comparison with subsequent p-value adjustment was required.

In order to investigate whether the E1/E2 metabolism in mammary gland tissue was decisive for the E1 and E2 mammary gland tissue levels and to identify factors influencing DNA adduct formation fluxes, linear regression analysis was used (Chapter 6.6.3). Other benefits of liner regression analysis are the quantification of the strength of relationships and identification of variables with redundant influence of the tested variables.

All statistical analysis were carried out with support of Prof. Ickstadt (Chair of Mathematical Statistics with Applications in Biometrics, University of Dortmund) and co-workers, especially Claudia Köllman and Helene König.

6.6.1 Correlation analysis of transcripts

In order to investigate associations of transcripts of enzymes involved in E1/E2 metabolism (Chapter 2.2.2) as well as in proliferation, cell cycle and apoptosis (Chapter 2.4) with continuous characteristics associated with breast cancer risk, including age and BMI (Chapter 6.1.1) and the mammary gland tissue level of the two main estrogens in mammary gland (E1 and E2, Chapter 5.2.14), correlation analysis was performed (Chapter 5.2.15). Since it was unknown how both E1 and E2 trigger the response in mammary gland tissue, the sum and the ratio of E2/E1 were also included in analyses.

Transcripts which were not determined in any or in most samples (UGT1A3/4 and SULT2A1, Chapter 6.3) and the reference genes GUSB and TBP were excluded from correlation analysis. Since the levels of some transcripts varied between samples by up to approx. 290% (Figure 28), the correlation analysis was carried out using Spearman correlation coefficient. The correlation coefficients were displayed in a colored and clustered heatmap (Figure 43).

In correlation analysis with 25 samples 122 associations with a correlation coefficient $\geq$0.5 were observed (Table A68). Two main clusters were observed (Figure 43) with one cluster including E1, E2, the ratio and the sum of E2 and E1 as well as three of the four marker-transcripts for ESR1 activation. The second cluster included most of the transcripts encoding E1/E2 metabolizing enzymes as well as transcripts involved in apoptosis and negative regulation of the cell cycle (Figure 43).

Correlation between E1/E2 and estrogen responsive genes

In order to investigate the association of the E1 and E2 tissue levels with the transcript levels of estrogen responsive genes and to investigate if the E1 and E2 tissue levels were correlated with each other, the Spearman correlation coefficient was used. Tissue levels of E2 and E1 (R=0.67), E2 (E1) and the sum of both (R=0.89 and 0.91, respectively, Figure 44), as well as E2 and the ratio E2/E1 (R=0.66) were highly correlated. Interestingly, the sum of E2+E1 and ESR1 (R=-0.55) were negatively correlated. A reduction of mRNA of ESR1 was observed in MCF-7 cells cultured with 10 nM E2 for 6 h (Stoica et al., 2003), but no data concerning normal mammary gland tissue has been published until now.

Ligand-activated ESR1 directly induces the transcription of TFF1, PGR and AREG (Chapter 2.2.2). Transcript levels of PGR were slightly associated with tissue levels of E2 and E1 (R=0.53 and 0.46, respectively) and with the sum of E2 and E1 (R=0.53, Figure 44). Furthermore, a strong correlation between transcript levels of the estrogen responsive genes TFF1 and AREG was observed (R=0.80, Figure 44). Neither the transcript level of TFF1 nor the transcript level of AREG were associated with the transcript level of PGR or E1/E2 tissue level.

Proliferation, differentiation and involution

Both CDKN1A and CDKN1B are involved in negative progression of cell cycle, while MKI67 and CCND1 are a markers for proliferating GECs (Chapter 2.1). As expected, the transcript levels of CDKN1A and CDKN1B correlated positively (R=0.62, Figure 44), while the transcript levels of CCND1 and MKI67 were not correlated (Figure 44). In addition, the transcript level of WNT4, involved in positive modulation of ductal side branching and extension, was positively correlated with the transcript level of MKI67 (R=0.58, Figure 44).

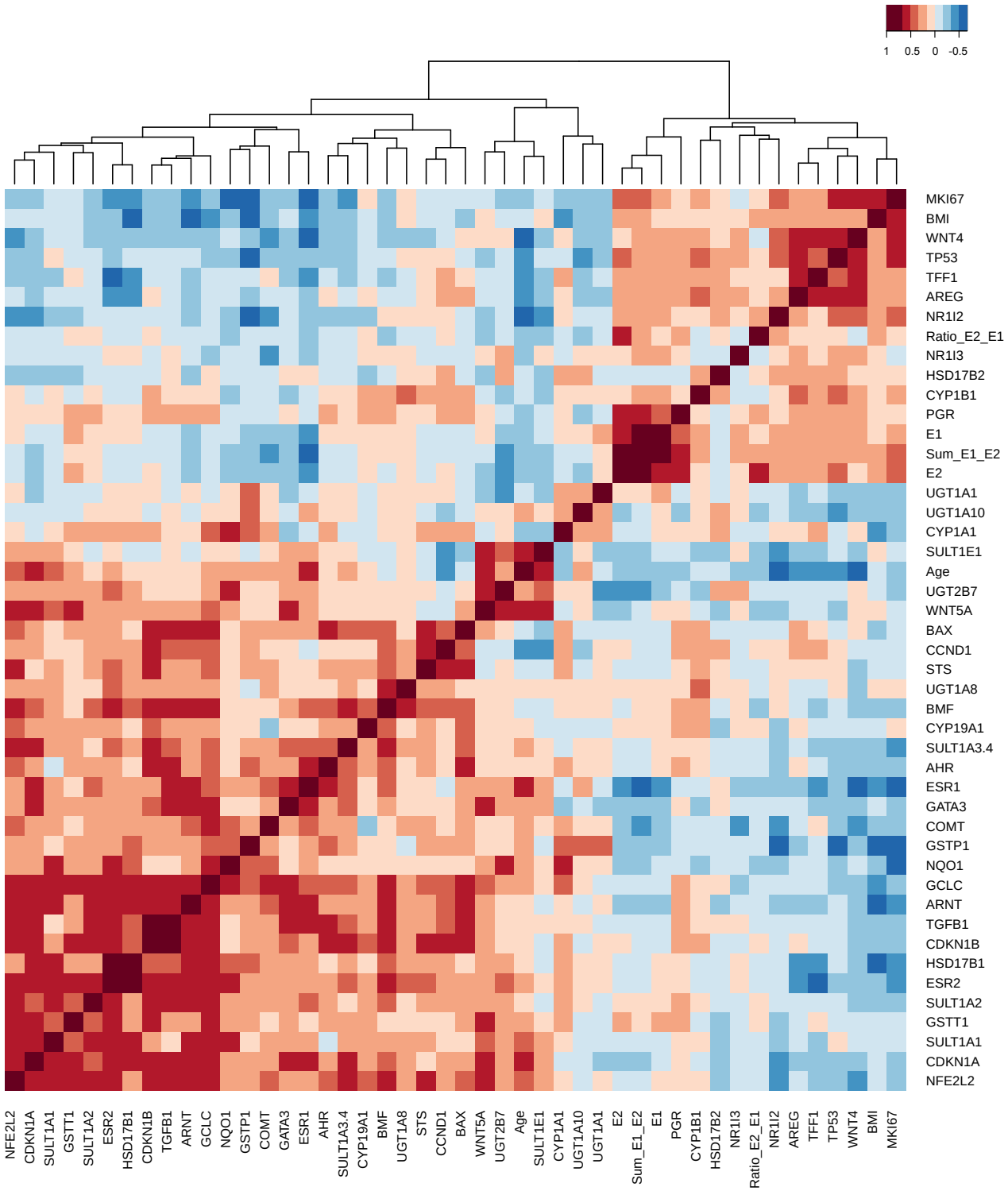


Figure 43: Clustered heatmap-representation of Spearman-correlation coefficients. Correlation analysis of transcript levels, age, BMI, E2 and E1 tissue level, the sum of E1 and E2 and the ratio E2/E1 was carried out with data of 25 mammary gland tissue samples.

A key inducer of GEC differentiation seemed to be GATA3 and WNT4, a downstream effector of PGR. WNT5A is a downstream effector of TGFB1 which is also involved in GEC differentiation. Thus both WNTs are involved in either positive (WNT4) or negative (WNT5A) modulation of ductal side branching and extension during mammary gland development (Chapter 2.1).

No correlations between both WNTs were observed, but both were clustered in each of the main clusters together with their respective upstream signal proteins: WNT5 together with TGFB1 and WNT4 together with PGR (Figure 43). The transcript levels of GATA3 (GEC differentiation) were also positively correlated with WNT5A (R=0.61) and TGFB (R=0.58, Figure 44). Furthermore, a very strong correlation was observed between the transcript levels of TGFB1 and CDKN1B (R=0.83), which was plausible, since both are involved in negative modulation of cell cycle (Chapter 2.1).

The proapoptotic BMF as well as BAX are also targeted by TGFB1 dependent signals and both are potential mediators of mammary gland involution (Chapter 2.2.2). The transcript levels of BMF and BAX were both positively correlated with TGFB1 (R=0.68 and R=0.65, respectively). Furthermore, both were positively correlated with the transcript level of CDKN1B (R=0.69 and R=0.78, respectively), which is also involved in cell cycle arrest signals. The transcription levels of BAX and BMF were also slightly associated (R=0.48, Figure 44).

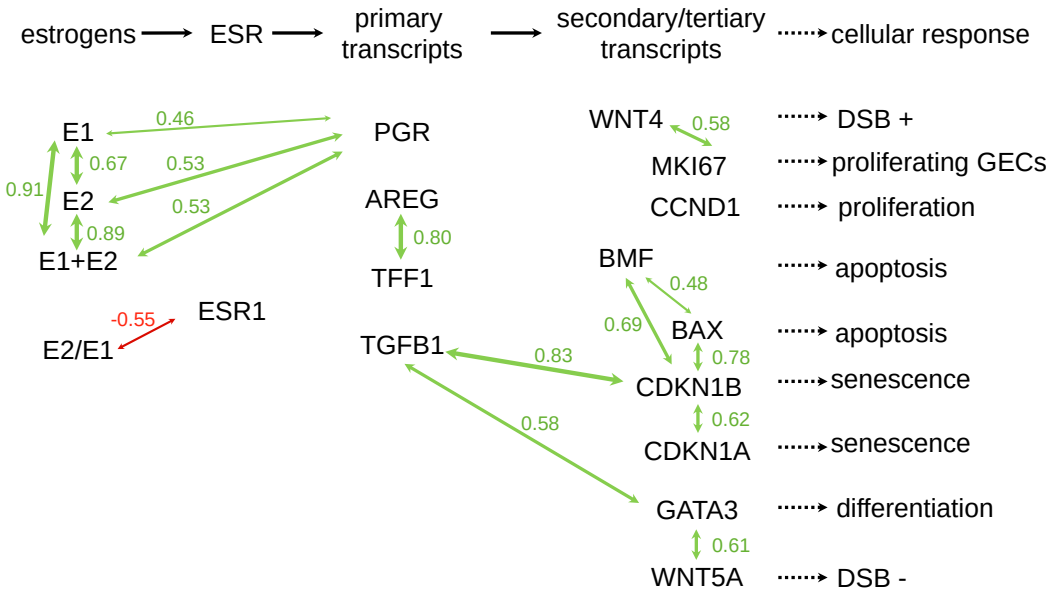


Figure 44: Overview of correlation coefficients between transcripts involved in proliferation, differentiation, apoptosis and cell cycle. DSB, ductal side branching.

E1/E2 metabolism

Although most of the isoenzymes of SULTs, UGTs, GSTs or CYPs are controlled by the

same transcription factors (Chapter 2.2.3), no correlations between isoenzymes were observed. A weak correlation between the transcript levels of NQO1, involved in detoxification in quinones, and CYP1A1 (Figure 43, R=0.55), the expression of both supposed to be controlled by AHR and ARNT (Chapter 2.2.3), was observed. The transcript of COMT, an ERE containing gene and therefore inducible by ESR1 activation, of which a reduction in transcript level due to E2 treatment in MCF-7 cells has been observed (Wagner et al., 2008), was slightly correlated in mammary gland tissue with sum of E1 and E2 tissue levels (Figure 43, R=-0.43).

In addition, the transcript levels of enzymes involved in E1/E2 metabolism with an ARE in their promotor region (GSTs, NQO1, GCLC, Chapter 2.3) were correlated with each other: GCLC-NQO1 (R=0.57), GCLC-GSTP1 (R=0.47), GCLC-GSTT1 (R=0.58), NQO1-GSTT1 (0.37), NQO1-GSTP1 (R=0.47), GSTT1-GSTP1 (R=0.25).

None of the correlations described above in normal mammary gland tissue has been published until now, but based on current knowledge summarized in Chapters 2.2.2 and 2.3 the described correlations were plausible. It must be considered that more associations of transcripts could not be excluded, since the interpretation of correlation coefficients is not uniform in literature and strongly dependent on samples size.

6.6.2 Influence of individual factors on transcript levels

Since factors correlated with breast cancer risk, such as BMI and smoking, could interact with E1/E2 metabolism, the aim was to investigate the associations between transcript levels of genes encoding enzymes involved in E1/E2 metabolism and the individual factors: age, BMI and smoking habit, as well as the lobule type (Chapter 2.1).

It must be considered that each group of the four individual factors contained different numbers of samples: Only seven of the 25 donors were characterized as lobule type 1_{parous} and six of the 25 donors were smokers (Table 47). Furthermore, twelve donors had a BMI <25, and six of them were smokers, which were all smokers in this study (Table 47). Thus, effects due to BMI could not be distinguished between effects due to BMI or smoking in the following statistical analyses. The impact of HAD was not investigated due to small sample sizes (n<3) after grouping HAD according to their composition (ethinylestradiol vs. estradiolvaleriate vs. progestin) and the possible impact on estrogen levels. Furthermore, some female donors provided insufficient information concerning the HAD used.

To investigate the associations of transcript level with age and BMI a correlation analysis of continuous variables was performed, as described in Chapter 6.6.1. The associations of the categorical lobule type and smoking habit were investigated by means of two-sample comparison and subsequent p value adjustment.

Table 47: Distribution of samples analyzed in this study into two categories of each characteristic. LOB, lobule type. n, absolute number of samples.

Characteristic	Category 1 (n)	Category 2 (n)
LOB	1 or 2/3 (18)	1_{parous} (7)
Age	<40 (11)	>40 (14)
BMI	<25 (12)	>25 (13)
Smoking	no (19)	yes (6)

6.6.2.1 Associations of lobule type and age

Age and lobule type are closely associated, since the lobule type changes with age in parous women (Chapter 2.1). The lobule type as categorical characteristic was investigated with two-sample comparison (Chapter 5.2.15), but no significant difference in any transcript level and E1/E2 tissue level (including the sum and the ratio of both) was observed (Table A69).

Age was negatively correlated with transcript levels of WNT4 (R=-0.51), NR1I2 (R=-0.63) and positively correlated with ESR1 (R=0.53), SULT1E1 (R=0.71), WNT5A (R=0.57) and CDKN1A (R=0.57, Figure 43).

In order to discriminate between effects based on age or on lobule type, the age was additionally categorized concerning expected age-related changes in lobule type: Since the lobule type of parous women is sharply decreasing back to lobule type 1_{parous} starting with age of 40 and the mean age for menopause is 51 (Russo and Russo, 2004), three categories were defined: age <40, between 40 and 50 and >50. Since only the trend of the values among the three different groups were of interest, a statistical evaluation was omitted. If the effect is age-related a continually increase/decrease over all three groups was expected. In contrast, if the effect was lobule type dependent, the increase/decrease would be sharply in the group of age 40-50 and then no further change in group of age >50 would be expected.

The transcript levels of NR1I2 were decreased in the group of donors at the age of >50 and 40-50 compared to the group of donors <40 (Figure 45). The transcript levels of WNT4 were decreased in the group of donors >50, both indicating possible impact of lobule type. A gradual increase with age was observed in transcript levels of ESR1, WNT5A, CDKN1A and SULT1E1 (Figure 45) indicating no effect of lobule type. Although all transcripts were correlated with age (Figure 43), changes in levels occurred at different ages, indicating a lobule type dependent change in transcript levels of NR1I2 and possibly WNT4.

In liver and small intestine of humans of different age and gender a decrease in transcript level of NR1I2 (PXR receptor, involved in expression various E1/E2 metabolizing enzymes) with age was found, whereas in lung and kidney no age related effect was observed (Miki et al., 2005). So, this effect seems tissue depended and has not been described in mammary

gland until now. Moreover, until now, mRNA of NR1I2 was only detected in breast cancer and adjacent normal tissue (summarized in Qiao et al., 2013), which is different to normal tissue of healthy women.

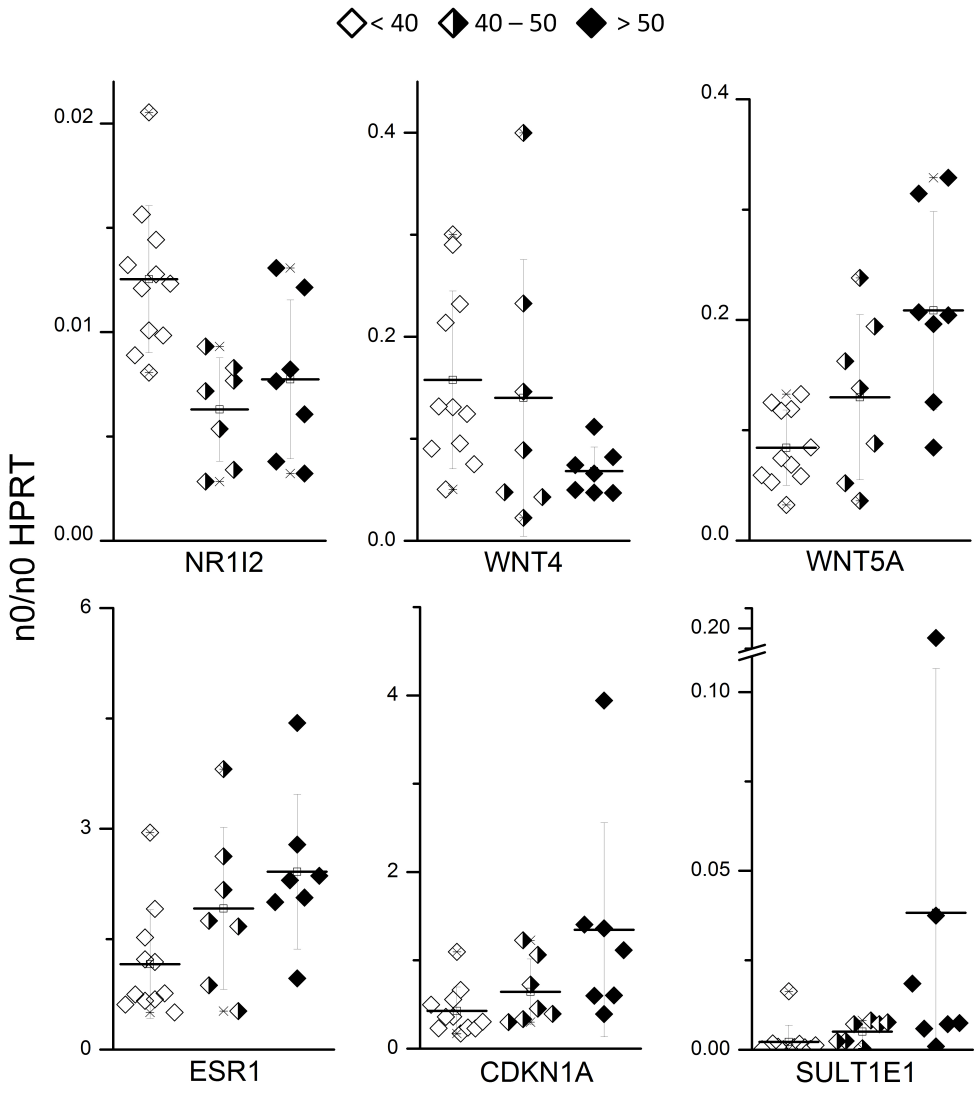


Figure 45: Transcript levels of NR1I2, WNT4, WNT5A, ESR1, CDKN1A and SULT1E1 in mammary gland homogenate (Chapter 6.3) divided into groups of different ages of the donors. No statistical evaluation was performed, since only the trend of change among the three age groups was of interest. Diamonds, samples. Horizontal line, mean. Vertical lines, SD.

The increase of mRNA levels of ESR1 and SULT1E1 with age were previously observed in tumor (Anders et al., 2008; Pfeiler et al., 2009) and in normal mammary gland tissue (Hallberg et al., 2008). Likewise, an increase in protein level of CDKN1A and a decrease in protein level of CDKN1B in normal breast tissue of postmenopausal women compared to premenopausal women was previously observed (Eigeliene et al., 2008a), a possible correlation

with age was not further investigated.

Both WNT4 and WNT5a are involved in ductal side branching and extension during mammary gland development (Chapter 2.1) and it is plausible that WNT4 (inducing side branching) is decreasing with age and WNT5A (inhibiting side branching) is increasing with age. A correlation of the WNT signaling pathway in normal mammary gland with age has been published based on microarray data using gene ontology analysis (Pirone et al., 2012). However, no transcript of the WNT signaling pathway was described in particular in this publication.

6.6.2.2 Associations of BMI and smoking habits with transcript levels

Both obesity and smoking are factors associated with a higher breast cancer risk (Chapter 2.4). Since all smokers (six samples) in this study had a BMI <25 (normal weight) between effects due to BMI or smoking could not be distinguished in the following statistical analyses. BMI correlated negatively with the transcripts of ARNT(R=-0.52), HSD17B1(R=-0.56) and GSTP1 (R=-0,71) and positively with MKI67 (R=0.63, Figure 43).

BMI

BMI is correlated with a higher risk for breast cancer in postmenopausal women. As one reason higher estrogen levels in blood of obese women, which could lead to higher mammary gland tissue concentration and thus higher proliferation, are discussed (Chapter 2.4). Thus, positive correlations with E1/E2 levels, transcripts controlled by ESR1 (TFF1, PGR, AREG), and transcripts expressed in proliferating cells (MKI67, CCND1, Chapter 2.1) were theoretically possible, but have not been published until now. Yet, no significant correlation between the BMI and E2 (Figure 43, R=0.28) and E1 mammary gland tissue level, as well as sum and ratio of E2/E1, was observed. Accordingly, no associations of BMI with transcripts related to ESR1 activation or proliferation were observed. The number of postmenopausal women in the study was too small to get meaningful results from a correlation analysis.

The transcript levels of HSD17B1 (R=-0.56) in mammary gland tissue homogenate decreased with increasing BMI, which may lead to less local conversion of E1 to E2 in mammary gland tissue, but no correlation of BMI and ratio E2/E1 tissue levels were observed (Figure 43). The influence of an reduced transcription level of ARNT (R=-0.52) in normal mammary gland tissue needs to be investigated. No data in normal mammary gland tissue concerning the correlation of ARNT and HSD17B1 have been published until now.

Smoking

Smoking was positively associated with transcript levels of GSTP1 (p=0.0083, Table A69). GSTs are involved in detoxification of reactive metabolites in cells, e.g. estrogen quinones, which are known to damage DNA (Chapter 2.2.2). The GSTP1 gene contains an ARE in

its promotor region and is therefore inducible by oxidative stress (Chapter 2.2.2). Thus, higher transcript level of GSTP1 in mammary gland tissue indicated cellular reaction in response to oxidative stress and it is known that smokers have higher levels of reactive oxygen species (Filaire et al., 2013). However, no tissue levels of markers for oxidative stress have been published until now, although, elevated levels of proinflammatory mediators were detected in adipose tissue of obese women (Chapter 2.4). The inflammation process is known to generate reactive oxygen species, known to damage DNA, which is also discussed as a mechanism for higher breast cancer risk of obese women (Chapter 2.4). Unfortunately, the associations of GSTP1 with BMI and smoking habit could not be ascribed to either BMI or smoking, therefore the analysis of more individuals is needed for clarification.

Further associations of transcripts could not be excluded, because for weak correlations the sample size might not be sufficient. In contrast, correlation coefficients of approx. 0.8 indicated a rather strong association (Helene König, Mathematical Statistics with Applications in Biometrics, University of Dortmund, personal communication) and the probability, that this effect is confirmed, is even higher.

However, the correlation coefficients only consider the two variables correlated, overlapping and counteracting variables are not taken into account. Therefore, the linear regression model allows a more sophisticated analysis of factors (e.g. transcript patterns) contributing to a variable (e.g. E2 tissue levels, Fahrmeir et al. 2013).

6.6.3 Linear regression models

In order to investigate if the metabolism in the mammary gland is decisive for E1 and E2 mammary gland tissue level, a statistical test was needed which could deal with more than one explanatory variables and rank them by strength of their influence. Thus, the linear regression model (Chapter 6.6) was chosen. The linear regression model was also used to identify variables affecting DNA adduct formation fluxes (calculated in Chapter 6.5.1.5) in normal mammary gland tissue.

To perform a linear regression analysis only complete data sets can be used. Furthermore, linear regression analysis requires more samples than variables. The resulting models for the dependent variables E2 and E1 tissue levels and ratio of E2/E1 were then optimized concerning a better (lower) Aikake information criterion (AIC, Chapter 5.2.17). This was done with a stepwise backward selection of the model (Chapter 5.2.17). Besides AIC, the adjusted R^2 is another marker for the quality of the model. No minimum value for the adjusted R^2 is described in literature, in general the higher the adjusted R^2 (maximum 1) the better the model describes the dependent variable.

6.6.3.1 Factors influencing E1 and E2 tissue level

Out of the data set 18 samples with complete data in E2 and E1 tissue levels (Chapter 5.2.14), in transcripts of enzymes directly metabolizing E1 and E2 and in individual factors (BMI, age, parity, smoking habit, HAD, lobule type, Chapter 6.1.1) were identified. E1 and E2 tissue levels and the ratio of both (E2/E1) were chosen as dependent variables to be analyzed with a linear regression model each to directly investigate differences in the influencing factors of E1 and E2 tissue levels. Since it is not precluded that factors associated with breast cancer risk affect both, E1 and E2 tissue levels, the six individual factors were also included in all linear models. Since the sum of E2+E1 tissue level was strongly correlated with E2 (R=0.89) and E1 (R=0.91, Chapter 6.6.1), it was not analyzed in the linear regression analysis. E1, E2 tissue level and the sum of E2+E1 were logarithmized, since the data was not normal distributed.

As explanatory variables, all transcripts of enzymes directly metabolizing E1 or E2 (CYP1A1, CYP1B1, UGT1A1, UGT1A8, UGT2B7, SULT1A1, SULT1A2, SULT1E1, STS, CYP19A1, HSD17B1 and HSD17B2, Chapter 2.2) and the individual factors possibly influencing E1 or E2 tissue level (BMI, age, parity, smoking habit, HAD, lobule type) were taken into account. Since only 18 samples with complete data set were available and the maximum number of variables tested in the linear regression model was 17, 18 variables would exceed the maximum number of variables for the regression analysis.

In order to reduce the variables, a principle component analysis (PCA) was performed with non-logarithmized transcripts encoding E1 and E2 metabolizing enzymes: Three principal components were identified (Figure A78) and used subsequently as explanatory variables in linear regression. Since E2 and E1 are metabolized by the same isoenzymes with the exception of SULT1A1, SULT1A2 and UGT2B7 which only metabolize E2, the linear regression analysis of both E1 and E2 tissue levels were performed with the three PCs (Table 48).

The first principle component (PC1) was mainly characterized by low transcript levels of SULT1A2 followed by HSD17B1, CYP1A1, STS and CYP19A1 (decreasing impact on PC1, Table 48). PC2 was mainly characterized by low transcript levels of UGT2B7, SULT1A1, HSD17B2 and high transcript levels of UGT1A1 and SULT1E1. PC3 was characterized by high transcript levels of UGT1A8, CYP1B1, UGT1A1 and STS.

Table 48: Description of the three PCs of transcripts of E1/E2 metabolizing enzymes. The higher the absolute value the bigger the influence on the PC (most influencing values in bold). The algebraic sign indicates the direction of influence: -, low transcript level; +, high transcript level.

Transcript	PC1	PC2	PC3
CYP1A1	**-0.48**	0.11	0.02
CYP1B1	0.02	-0.19	**0.50**
UGT1A1	-0.02	**0.37**	**0.31**
UGT1A8	-0.02	-0.06	**0.61**
UGT2B7	0.04	**-0.53**	-0.05
SULT1A1	-0.09	**-0.46**	-0.17
SULT1A2	**-0.50**	-0.08	-0.11
SULT1E1	0.04	**0.25**	0.06
STS	**-0.40**	-0.02	**0.39**
CYP19A1	**-0.29**	-0.14	-0.12
HSD17B1	**-0.49**	-0.03	-0.07
HSD17B2	0.15	**-0.48**	0.25

E2-model

The E2 tissue level was significantly negatively correlated with the lobule type 1_{parous} (p=0.0033) and PC3 (p=0.0365, Table 49).

This indicated, that in mammary gland tissues of the lobule type 1_{parous} the E2 tissue levels were decreased compared to the tissue level of lobule type 2/3. A low E2 plasma level is described in postmenopausal women (Blair, 2010). Since the lobule type 1_{parous} is observed in parous women in their fifth decade, the postmenopausal status (mean age of menopause=51, Russo and Russo, 2004) is roughly comparable (Chapter 2.2.1). However, the decrease of the tissue E2 level in postmenopausal mammary gland tissue has been only described in cancer tissue and normal tissue adjacent to cancer tissue until now (Lonning et al., 2009).

The higher PC3 (higher trancript levels of UGT1A8, CYP1B1, UGT1A1 and STS) the lower the E2 tissue level. This is plausible for UGT1A8, CYP1B1 and UGT1A1, since the enzymes are metabolizing E2. For the transcript of STS, this is not plausible, since reaction catalyzed by STS results in free E2 (Chapter 2.2.2). The transcript level of STS, however, had the least influence, of the four transcript level, on the PC3. Thus, this result needed to be confirmed with a linear regression model including more samples to be able to consider transcript levels instead of PCs.

E1-model

The E1 tissue level was negatively related to parity (nulliparous, p=0.0147) and non-smoking (p=0.0297, Table 49). This indicated that nulliparous women had higher E1 mammary gland tissue levels, and smokers lower E1 tissue levels. Until now, no association between the E1 tissue levels and parity had been described. A study with healthy women showed no

association between the urinary E1 levels and parity (Fortner et al., 2012).

A lower serum E1 level has been related to smoking by nicotine increasing the number of regressing follicles in the ovary and inhibiting CYP19A1, which leads to a decreased conversion of androgens to estrogens (summarized in Tanko and Christiansen, 2004), but the effect on mammary gland tissue has not been described until now. However, it must be considered, that the adjusted R^2 of this model (R^2=0.2065) was rather low.

E2/E1-model

The ratio of E2/E1 was strongly correlated (negatively) to the lobule type 1_{parous} (p=0.0009) as well as to PC2 (p=0.0143) and PC3 (p=0.0112, Table 49).

This indicated a lower E2/E1 ratio in mammary gland tissue of women with lobule type 1_{parous}.

PC3 was mainly influenced by UGT1A8 and CYP1B1. The K_m values for glucuronidation of E1 and E2 by UGT1A8 are slightly different (55 and 38 μM, respectively Lepine et al., 2004). Furthermore, the hydroxylation to respective catechol estrogens by CYP1B1, specified as nmol catechol estrogen/min/nmol CYP1B1, was less for E1 than for E2 (e.g. formation of 4HO-E1/E2: 0.24 and 0.91 nmol/min/nmol CYP1B1, respectively, Shimada et al., 1999). Hence, both reactions would result in higher E1 levels.

The ratio of E2/E1 was also significantly negatively correlated with PC2, which is mainly influenced by low transcript levels of UGT2B7, which only catalyzes the glucuronidation of E2 and HSD17B2, which converts E2 to E1. Thus, both enzymes catalyze reactions which result in higher E1 than E2 level. In sum, these results indicate that the different affinity of enzymes either to E1 or E2 is reflected in the respective mammary gland tissue level. However, to elucidate the roles of these enzymes on E2/E1 ratio in tissue level in detail, more samples are needed to preform a linear regression model without PCs, which are always associated with a loss of information.

Table 49: Final linear regression models of E2, E1 and E2/E1 performed with PCs of the transcripts of E1 and E2 metabolizing enzymes and six confounding factors (BMI, age, parity, smoking habit, HAD, lobule type). The presented factors were included in the final model after stepwise backward selection. Statistical significance is indicated by p value<0.05. Model (M) E2: adjusted R^2=0.4759. M-E1: adjusted R^2=0.2065. M-E2/E1: adjusted R^2=0.5783. $LOB1_{parous}$, samples with lobule type 1_{parous}.

M	Factor	p value		M	Factor	p value		M	Factor	p value	
E2	Parity	0.0500		**E1**	Parity	0.0284	*	**E2/E1**	Parity	0.0720	
	Smoking	0.1251			Smoking	0.0297	*		Smoking	0.0663	
	HAD	0.2349			$LOB1_{parous}$	0.1701			$LOB1_{parous}$	0.0009	***
	$LOB1_{parous}$	0.0033	**						PC2	0.0143	*
	PC2	0.1266							PC3	0.0112	*
	PC3	0.0365	*								

6.6.3.2 Factors influencing DNA adduct formation

In order to investigate the association of the calculated DNA adduct formation fluxes and factors influencing breast cancer risk, a linear model was set up including all individual factors (BMI, age, parity, smoking habit, HAD, lobule type, genotype of COMT Val108/158Met, NQO Pro187Ser and GSTT1 CNP) as well as the tissue levels of E1, E2 and the ratio E2/E1 (Chapter 6.1), since no reasonable pre-selection of factors could be made based on previously published data. Twenty-two complete observations were identified and used for the calculation of the linear model.

The final model included E1, the ration E2/E1, parity, age, BMI, smoking habits, the use of HAD and both polymorphisms COMT Val108/158Met and GSTT1 CNP (Table 50).

The most influencing factor on DNA adduct formation flux was the use of HAD. Donors using HAD had lower DNA adduct formation fluxes (p=0.004, Table 50). It must be taken into account, that the factor HAD in this model was used without considering the different composition of HADs, thus a discussion of this result is difficult. More samples are needed to divide the use of HAD into appropriated subgroups, e.g. differentiating between preparations containing E2 or Ethinylestradiol and administration form, local or oral. As indicated earlier (Chapter 6.6.2) the group of HAD-users was very small, so that a subdivision was not appropriate. Additionally, some donors provided insufficient information concerning the HAD used. Hence, at this point it could only be stated, that there is an influence on the calculated DNA adduct forming fluxes, but to assign this result to a specific effect more data is needed.

BMI and GSTT1 CNP were associated positively with DNA adduct formation flux:

Donors with higher BMI had higher DNA adduct formation fluxes (p=0.014). Assuming increased calculated DNA adduct fluxes lead to more DNA adducts in the cells, the number of mutations may also increase, which in turn may lead to an increased number of initiated cells and thus to a higher breast cancer risk (Figure 8). These assumptions were made without considering important factors such as apoptosis and DNA repair, which may prevent the cells from DNA damage.

Genotypes of GSTT1 +/+ and +/- were negatively correlated with DNA adduct forming fluxes (p=0.014 and p=0.020, respectively, Table 50). The GSTT1 -/- is also correlated with a higher breast cancer risk (Chapter 2.4.1.2). Thus, under the assumptions made above, this result was also in accordance to epidemiology. However, it must be taken into account, that only two of the 22 samples were GSTT -/-, so this result remains to be proven with a larger sample group with genotype GSTT1 -/-.

Both, age and parity were weakly associated with DNA adduct formation fluxes:

DNA adduct fluxes calculated with transcript levels of younger donors were higher than

that of older donors (p=0.046). Likewise, nulliparous donors had higher fluxes (p=0.064, Table 50), although breast cancer is more frequent in older than in younger women (Chapter 2.4). Due to the slow, stepwise process of carcinogenesis, at the time point of diagnosis, tumors are the results of mutations initiated up to decades ago.

Nulliparous women are known to have a higher risk of breast cancer (Chapter 2.4), which was also reflected in the DNA adduct formation fluxes calculated in this work.

Strikingly, E2 mammary gland tissue levels, suspected to be an important factor affecting breast cancer risk, were not decisive for DNA adduct fluxes. This could be due to the fact, that the level of E1 and E2 were set as unlimited available in the network-based calculation of DNA adduct forming fluxes.

Furthermore, the polymorphisms of NQO1 Pro187Ser were not decisive for the calculated DNA adduct forming fluxes, although in epidemiological studies the risk of breast cancer for the NQO1 Ser/Ser genotype was elevated by 27%, which was even higher than the increased risk for the GSTT -/- genotype (11%, Chapter 2.4.1). There are three possible reasons for this: (i) The detoxification of reactive estrogen metabolites by NQO1 is not the only reaction which is responsible for the elevated risk in epidemiological studies, since NQO1 also detoxifies other reactive quinones. (ii) The statistical model: If there were two redundant influnces of two explanatory variables, one is eleminated from the final model. Thus, the influence of the NQO1 polymorphism on the calculated DNA adduct forming fluxes could be eleminated from the model due to redundancy. (iii) Since there were no NQO1 Ser/Ser genotyped samples in this study, the effect of NQO1 Pro/Ser on the DNA adduct forming flux might be too weak, although the NQO1 Pro/Ser genotypes was also correlated with a higher breast cancer risk in epidemiological studies (Chapter 2.4.1).

In sum, it must be taken into consideration that some effects may be biased by settings made in network calculation (Chapter 6.5.1.6) and results could also be biased due to the chosen statistical analysis. Furthermore, the comparison to epidemiological studies was made using the calculated DNA adduct flux as a marker for breast cancer risk, regardless of other important factors in carcinogenesis (Chapter 2.3.1) such as proliferation, apoptosis and DNA repair.

Despite this, the calculated DNA adduct formation fluxes of the 22 samples used for this regression model reflected most of the results of epidemiological studies concerning breast cancer risk. This would support the assumption, that the metabolism of E1 and E2 to reactive metabolites plays an important role in development of breast cancer and is directly affected by BMI, GSTT CNP, age and parity.

Table 50: Final linear regression model of DNA forming adduct flux calculated in this work (Chapter 6.5.1) performed with BMI, age, parity, smoking habit, HAD, lobule type, genotype and tissue levels of E1, E2 and the ratio E2/E1. The presented factors were included in the final model after stepwise backward selection. Statistical significance is indicated by p value$<$0.05, adjusted R^2=0.7691.

Factor	p value	
E1	0.261	
E2/E1	0.226	
Parity	0.064	
Age	0.046	*
BMI	0.014	*
Smoking	0.251	
HAD	0.004	**
COMT Val/Met	0.891	
COMT Val/Val	0.101	
GSTT1 +/-	0.014	*
GSTT1 +/+	0.020	*

7 Summary

The female sex hormone 17β-estradiol, produced naturally in the body, seems to play an important role in the development of breast cancer, since (i) it can be activated to reactive metabolites, which are known to damage DNA and (ii) the stimulation of the estrogen receptor α by 17β-estradiol enhances cell proliferation. Both processes together increase mutation frequency and subsequently lead to transformation of epithelial cells. Therefore, the aim of this work was to characterize the influence of polymorphisms and lifestyle factors on 17β-estradiol metabolism in normal mammary gland tissue.

As important lifestyle factors associated with breast cancer risk by presumably modulating 17β-estradiol metabolism age, BMI -especially for postmenopausal women-, smoking and number of pregnancies were identified. It was focused on polymorphisms in genes encoding for enzymes involved in detoxification of reactive metabolites, namely, glutathione-S-transferase theta 1 and NADPH-quinone oxidoreductase 1 involved in detoxification of quniones and catechol-*O*-methyltransferase involved in detoxification of catechols.

Thirty normal mammary gland tissue samples were collected from women who underwent reduction mammaplasty for cosmetic reasons. The preparation of the mammary gland tissue for further analyses was standardized concerning maximum content of surrounding adipose tissue (25%) and time from removing the tissue until shock freezing of the tissue aliquotes in liquid nitrogen (15 minutes).

Since menopause seemed to be an important factor and this study was carried out on breast tissue, changes in breast morphology pre-dating menopause (lobule type) were histochemically classified, too.

The polymorphism in glutathione-S-transferase theta 1 directly influences the copy number of the gene (+/+, +/- or -/-) and thus indirectly enzyme activity. A previously published restriction fragment length PCR method was used to determine the genotype of the copy number polymorphism of glutathione-S-transferase theta 1. The single nucleotide polymorphisms in NADPH-quinone oxidoreductase 1 (Pro187Ser) and catechol-*O*-methyltransferase (Val108/158Met) result in reduction of enzyme activity due to shorter half-life or decreased thermostability of the protein, respectively. Both single nucleotide polymorphisms were determined using a commercially available multiplex TaqMan™-probe based allelic discrimination assay. As positive control for each genotype of the three polymorphisms, the genotypes of seven cultured human cell lines were identified. Furthermore, the allelic discrimination assay was verified by comparison to the restriction fragment length PCR, a well established method for the determination of single nucleotide polymorphisms.

As expected, the frequencies of the genotypes of all three polymorphisms matched the Hardy-Weinberg Equilibrium and were comparable to previously published frequencies

among the German population.

The 17β-estradiol metabolism involved 33 enzymes of which 26 were previously reported as present in human mammary gland. However, three of the non-detected enzymes were only investigated in one or two samples, namely uridine 5'-diphospho-glucuronosyltransferase 1A1 and 1A10 and sulfotransferase 2A1. Eight of the present transcripts were not considered in this work due to redundant or minor role in 17β-estradiol metabolism. Additionally, six transcripts of nuclear receptors involved in induction 17β-estradiol metabolizing enzymes and 19 transcripts for characterization e.g. of 17β-estradiol mediated proliferation, as well as three reference genes were selected.

The quantitative determination of the 49 transcripts was established using TaqMan™ low density array enabling the simultaneous quantification of 47 transcript levels of eight samples. The remaining two transcripts were quantified in an additional TaqMan™-probe based PCR in 96-well plates.

The variation of one triplicate within one slot of the array card were strongly dependent on the cycle threshold: the modal value of the relative standard deviations of 904 triplicates were 7.2% (cycle threshold 15-20), 6.7% (cycle threshold 20-25), 7.5% (cycle threshold 25-30) and 11.9% (cycle threshold 30-33). The relative standard deviation of cycle threshold values >33 were sharply increasing and therefore not considered for quatification. The intra-variation of the TaqMan™ low density array was comparable to inter-variation. The mean relative standard deviation of both ranged from 5% to 20% for 27 transcripts of the 31 quantifiable transcripts (cycle threshold range 15-33). The relative standard deviation of the remaining four transcripts ranged from 24% to 48%. In order to guarantee an unbiased quantification for less abundant transcripts, the linearity of the preamplification reaction of each transcript was invetestigated.

The preamplification reaction was linear for 44 of 49 transcripts. Since two of the non-preamplifiable transcripts encoded important enzymes of the 17β-estradiol metabolism (glutathione-S-transferase pi 1 and NADPH-quinone oxidoreductase 1) and the amounts of the transcripts in mammary gland tissue were sufficient for a determination without preamplification, both transcripts were quantified without preamplification. The remaining three transcripts, which were not preamplifiable, were not further investigated in this work.

Twenty-five of the 30 mammary gland tissue samples yielded enough RNA for quantification of low expressed transcripts using TaqMan™ low density array. The presence of two transcripts of the uridine 5'-diphospho-glucuronosyltransferase family, 1A1 and 1A10, in normal mammary gland tissue was revealed due to a larger number of samples investigated than in previous studies.

All transcript levels of enzymes involved in 17β-estradiol metabolism quantified in this work were higher in normal tissue than in MCF-7 cells. Moreover, three transcripts were

either not detectable or not quantifiable in cultured MCF-7 cells, while all three were determined in normal mammary gland tissue.

Additionally, transcripts involved in cell communication, like receptors (except for estrogen receptor α) and transcripts encoding enzymes of paracrine signal transduction, were more abundant in normal tissue than in MCF-7 cells. Overall, normal mammary gland tissue seemed to have a higher 17β-estradiol metabolism and associated signaling than MCF-7 cells.

A method for the isolation of epithelial cells from the surrounding stromal cells and adipose tissue was established using laser capture microdissection. The transcript levels of genes encoding enzymes involved in 17β-estradiol metabolism were quantified using a TaqMan™ low density array with double preamplification due to the very low amount of RNA. Thirteen transcripts, thereof ten transcripts involved in 17β-estradiol metabolism, were not detected in epithelial cells, all also had very low levels in homogenate. Thus, the investigation of 17β-estradiol metabolism in isolated epithelial cells was not possible. The comparison of the transcript levels of epithelial cells and homogenate involved in estrogen receptor α activation, signal transduction, proliferation and differentiation by cluster analysis revealed differences in cluster structure: The clusters of homogenate differentiated between lobule types, i.e. extend of morphologically visible mammary gland differentiation. In contrast, the clustering of epithelial cells differentiated between estrogen receptor α activation, high and low proliferation and differentiation. Thus, data derived from homogenate was used for further analysis of 17β-estradiol metabolism but analysis of glandular epithelial cell proliferation was not further pursued. For estimation of the influence of lifestyle and polymorphisms on spontaneous mutation frequency the Random Mutation Capture assay, published for the p53 gene in cultured cells, was applied to mammary gland tissue. No data on human normal mammary gland tissue has been published until now, thus for the first time

the spontaneous mutation frequency of two samples of normal mammary gland tissue were determined ($3.6\text{x}10^{-7}$ and $7.2\text{x}10^{-8}$). However, the original method was very laborious and time consuming. Hence, it was optimized concerning copy number determination of the target sequence and mutant determination: Two steps of gel electrophoresis were omitted and the reaction volume of the PCR for the mutant detection was reduced. Nevertheless, maintaining the sensitivity of the whole assay was extremely difficult and laborious. Therefore, the method was still not suitable for the amount of samples in this study.

Instead, a metabolic network containing 17β-estradiol metabolism was constructed, in co-operation with the department of Bioinformatics at the University of Würzburg, to calculate the DNA adduct formation flux as further marker for genotoxicity. For this purpose, 14 of the 142 total reactions were found in public databases, the remaining reactions were set up individually through literature search. Additionally, previously published network

reactions of co-factor production and energy metabolism (e.g. pathways of citric acid cycle and oxidative phosphorylation) were added to the metabolic network. In order to investigate the influence of individual differences in transcript levels of genes involved in 17β-estradiol metabolism and polymorphisms on genotoxicity, the DNA adduct formation fluxes were calculated for each sample based on the transcript levels quantified in this work. The DNA adduct formation fluxes were dramatically influenced (up to 9300% difference among samples) by the individual transcript level patterns of the 17β-estradiol metabolizing enzymes. In contrast, the polymorphisms had only a minor influence: The highest increase in DNA adduct formation flux (82.6%) was observed for a sample with a medium NADPH-quinone oxidoreductase 1 activity combined with low activity of catechol-*O*-methyltransferase compared to the respective DNA adduct formation flux without considering the genotype in the network calculation. Hence, the influence of the basal transcript level of each enzyme, possibly influenced by individual lifestyle, seemed to be stronger than the influence of polymorphisms on whole 17β-estradiol metabolism.

To answer the question if the factors influencing breast cancer risk, e.g. BMI and smoking, were associated with DNA adduct formation flux calculated on the basis of 17β-estradiol metabolism, linear regression analysis was used. A high BMI and the genotype -/- of glutathione-S-transferase theta 1 were highly correlated with higher DNA adduct formation fluxes. Age was negatively correlated with DNA adduct formation flux. Furthermore, nulliparous women had higher DNA adduct formation fluxes. Thus, the calculated DNA adduct formation fluxes reflected most of the results of epidemiological studies concerning breast cancer risk, thus adding mechanistic evidence to these descriptive observations.

A correlation analysis was performed to reduce the variables for linear regression analysis in the first place, but revealed no possibility for a reduction of variables. Interestingly, associations of continuous factors influencing the breast cancer risk (age and BMI) with transcripts quantified in this work were observed. Therefore, a two-sample comparison analysis with subsequent p-value adjustment was performed for the further investigation of the association of the non continuous variables (lobule type and smoking) with transcript levels. Already known correlations for age and BMI were detected, as well as new ones: Wingless related MMTV integration site-4 and 5A, both involved in ductal side-branching of the mammary gland, and the pregnane X receptor, involved in induction of various 17β-estradiol metabolizing enzymes, were correlated to age. Furthermore, BMI was positively correlated to the proliferation-related Ki-67 antigen and negatively to three transcripts of the 17β-estradiol metabolism. Smoking was significantly associated with the transcript level of glutathione-S-transferase pi 1, involved in detoxification of reactive quinones. Thus, age, BMI and smoking influenced transcript levels of enzymes involved in 17β-estradiol metabolism in normal mammary gland tissue.

Furthermore, for statistical investigation whether transcript levels of genes involved in 17β-estradiol metabolism are decisive for the 17β-estradiol mammary gland tissue level a linear regression analysis was used. Since this statistical model is limited in the number of variables and the correlation analysis did not suggest reduction of variables, the number of variables was reduced by performing a principal component analysis on the transcript levels of enzymes directly metabolizing 17β-estradiol. Three principal components were included together with individuals factors, e.g. BMI and age, in the statistical calculation. The lobule type, pre-dating menopause, had the strongest (negative) influence on 17β-estradiol tissue levels. Decreasing 17β-estradiol tissue level in menopause has been only described in cancer tissue and normal tissue adjacent to cancer tissue until now. Since the decreasing tissue level could be due to ceasing ovarian function, this result did not support the hypothesis that the tissue metabolism is decisive for the 17β-estradiol tissue level. Nevertheless, a significant influence of one principal component of the 17β-estradiol metabolizing transcript levels was detected, indicating the influence of tissue metabolism on the 17β-estradiol mammary gland tissue level. Thus, both the tissue metabolism as well as the plasma concentration of 17β-estradiol seemed to influence the 17β-estradiol tissue level. However, for a more detailed result, more samples were needed to omit the reduction of data and thus information by the principal component analysis.

In sum, the tissue specific 17β-estradiol metabolism was described in mammary gland tissue homogenate, whereas differences in proliferation of epithelial cells were only reflected in isolated epithelial cells. Factors associated with breast cancer risk (age, BMI and age-related changes in mammary gland morphology) were shown to affect 17β-estradiol tissue levels. The 17β-estradiol mediated genotoxicity was evaluated using bioinformatically calculated DNA adduct fluxes, which were predominately influenced by individual mRNA patterns rather than individual genotypes and (DNA adduct fluxes) were correlated with known breast cancer risk factors (age, parity, BMI and polymorphism of glutathione-S-transferase theta 1).

8 Zusammenfassung

Das körpereigene, weibliche Geschlechtshormon, 17β-Estradiol spielt eine wichtige Rolle bei der Brustkrebsentstehung, da (i) es zu reaktive Metaboliten aktiviert werden kann, welche die DNA schädigen können und (ii) durch die Stimulation des Estrogenrezeptors α die Zellproliferation steigern kann. Beide Prozesse können dann zum Anstieg der Mutationsfrequenz und anschließender maligner Transformation von Epithelzellen führen. Deshalb war das Ziel dieser Arbeit, den Einfluss von Polymorphismen und der Lebensweise auf den gewebespezifischen 17β-Estradiol-Metabolismus im normalen Brustdrüsengewebe zu untersuchen.

Als wichtige Faktoren, die mit einem erhöhten Brustkrebsrisiko assoziiert sind und die möglicherweise den 17β-Estradiol-Metabolismus beeinflussen, wurden das Alter, der BMI -vor allem bei postmenopausalen Frauen-, Rauchen und die Anzahl an Schwangerschaften identifiziert. Es wurden Polymorphismen in Genen ausgewählt, die für Enzyme kodieren, die die Detoxifizierung von reaktiven Metaboliten katalysieren: Die Glutathion-S-Transferase Theta 1 und NADPH-Chinon-Oxidoreduktase 1, die beide an der Entgiftung reaktiver Chinone beteiligt sind, und die Catechol-*O*-Methyltransferase, die die Entgiftung von Catecholen katalysiert.

Es wurden 30 Brustgewebsspenden von gesunden Frauen, die sich aus kosmetischen Gründen einer Mamma-Reduktionsplastik unterzogen haben, gesammelt. Die Präparation des Brustdrüsengewebes für die Untersuchungen wurde hinsichtlich des Höchstgehalts an umliegenden Fettgewebe (25%) und der Zeit von der Entfernung der Gewebe bis zum Schockgefrieren der Gewebe Aliquote in flüssigem Stickstoff (15 Minuten) standardisiert.

Da die Menopause als ein weiterer wichtiger Faktor im Zusammenhang mit der Brustkrebsentstehung erschien, wurden die Veränderungen in der Brustmorphologie, die kurz vor der Menopause auftreten (Lobulus-Typ), histochemisch klassifiziert.

Der Polymorphismus im Glutathion-S-Transferase Theta 1-Gen beeinflusst direkt die Kopienzahl des Gens (+/+, +/- oder -/-) und damit indirekt die Enzymaktivität. Eine zuvor veröffentlichte Methode, basierend auf einer Restriktionsfragmentlängen-Polymerasekettenreaktion, wurde verwendet, um den Genotyp des Glutathion-S-Transferase Theta 1-Polymorphismus zu bestimmen. Die Einzel-Nukleotid-Polymorphismen im NADPH-Chinon-Oxidoreduktase 1-Gen (Pro187Ser) und Catechol-*O*-Methyltransferase-Gen (Val108/158Met) führen zu einer Verringerung der Enzymaktivität, jeweils entweder durch eine kürzeren Halbwertszeit oder eine verringerte Thermostabilität des Proteins. Zur Bestimmung der Einzel-Nukleotid-Polymorphismen wurde eine kommerziell erhältlicher Multiplex TaqMan™-Sonden basierter Allelische Diskriminierungs-Assay verwendet. Sieben humane Zelllinien wurden genotypisiert, um für jeden Genotyp der drei Polymorphismen eine Positivkontrolle zu bestimmen. Weiterhin wurde die allelische Diskriminierung durch Vergleich mit der Restrik-

tionsfragmentlängen-PCR, eine etablierte Methode zur Bestimmung von Einzel-Nukleotid-Polymorphismen, verifiziert. Wie erwartet, entsprachen die Frequenzen der Genotypen aller drei Polymorphismen der 30 Brustdrüsengewebespenden dem Hardy-Weinberg Equilibrium und waren ebenfalls vergleichbar mit zuvor veröffentlichten Frequenzen in der deutschen Bevölkerung.

Am 17β-Estradiol-Metabolismus sind 33 Enzyme beteiligt, von denen 26 im menschlichen Brustdrüsengewebe exprimiert werden. Jedoch wurden in den Studien in denen drei (Uridin-5'-diphospho-glucuronosyltransferase 1A1 und 1A10, Sulfotransferase 2A1) der nicht-detektierten Enzyme untersucht wurden, nur eine oder zwei Proben untersucht. Acht der 26 Transkripte wurden in dieser Arbeit nicht Betracht gezogen, da sie eine redundante oder untergeordnete Rolle im 17β-Estradiol-Metabolismus der weiblichen Brustdrüse spielten. Zusätzlich zu den Transkripten des 17β-Estradiol-Metabolismus wurden sechs weitere Transkripte von Kernrezeptoren, die an der Induktion von 17β-Estradiol Enzyme beteiligt sind, und 19 Transkripte, zur Charakterisierung der 17β-Estradiol vermittelten Proliferation, sowie drei Referenzgenen ausgewählt.

Die quantitative Bestimmung der 49 Transkripte wurde mit TaqMan™-Sonden basierten Low Density Array durchgeführt, da es die gleichzeitige Quantifizierung von 47 Transkriptleveln von acht Proben ermöglicht. Die restlichen zwei Transkripte wurden in einer zusätzlichen TaqMan™-Sonden basierten PCR im 96-Well-Format quantifiziert. Die Schwankung von einer Dreifach-Bestimmung in einem Slot des Arrays waren stark vom Zyklusschwellenwert abhängig: der Modalwert der relativen Standardabweichungen von 904 Dreifachbestimmungen war 7,2% (Zyklusschwellenwert 15-20), 6,7% (Zyklusschwellenwert 20-25), 7,5% (Zyklusschwellenwert 25-30) und 11,9% (Zyklusschwellenwert 30-33). Die relative Standardabweichung ab einem Zyklusschwellenwerten über 33 war schnell ansteigend und daher wurden Transkripte mit einem Zyklusschwellenwert über 33 nicht quantifiziert. Die Intra- und Inter-Array-Schwankung des TaqMan™-Low-Density Arrays war vergleichbar. Die mittlere relative Standardabweichung lag zwischen 5% und 20% für 27 der 31 quantifizierbaren Transkripte (Zyklusschwellenwert 15-33). Die relative Standardabweichung der verbleibenden vier Transkripte reichte von 24% bis 48%. Um eine Verzerrung der Quantifizierung der Transkriptlevel zu vermeiden, wurde die Linearität der Prä-amplifizierungsreaktion jedes Transkripts überprüft. Die Prä-amplifizierungsreaktion war linear für 44 von 49 Transkripten. Zwei der nicht-prä-amplifizierbaren Transkripte kodieren wichtige Enzyme des 17β-Estradiol-Metabolismus (Glutathion-S-Transferase pi 1 und NADPH-Chinon-Oxidoreduktase 1). Da die Menge der Transkripte im Brustdrüsengewebe ausreichend war, wurden beide deshalb ohne Prä-amplifizierung quantifiziert. Die restlichen drei Transkripte, die nicht prä-amplifizierbar waren, wurden in dieser Arbeit nicht weiter untersucht.

Von 25 der 30 Brustdrüsengewebsspenden wurde genug RNA zur Quantifizierung von niedrig exprimierten Transkripten mittels TaqMan™-Low-Density Arrays isoliert. Das Vorhandensein von zwei Transkripten der Uridin-5'-diphospho-Glucuronosyltransferase Familie, 1A1 und 1A10 wurde aufgrund einer größeren Anzahl von untersuchten Proben normalen Brustdrüsengewebes nachgewiesen. Alle Transkriptlevel von Enzymen des 17β-Estradiol-Metabolismus waren in normalem Gewebe höher als in MCF-7-Zellen. Außerdem waren drei Transkripte in MCF-7-Zellen entweder nicht nachweisbar oder nicht quantifizierbar, während alle drei Transkripte in normalem Gewebe bestimmt wurden. Weiterhin wurden Transkripte, die an der Zellkommunikation beteiligt sind, wie beispielweise Rezeptoren (außer der Estrogenrezeptor α) und Enzyme der parakrinen Signalweiterleitung in größeren Mengen in normalen Gewebe, als in MCF-7 Zellen quantifiziert. Insgesamt schien normales Brustdrüsengewebe einen höhere 17β-Estradiol-Metabolismus und die dazugehörigen Signalweiterleitung zu haben, als MCF-7-Zellen.

Eine Methode zur Isolierung von Epithelzellen aus den umliegenden Stroma- und Fettzellen des Brustdrüsengewebes mittels Laser-Mikrodissektion wurde etabliert. Die Transkriptlevel der Gene, die für Enzyme des 17β-Estradiol-Metabolismus kodieren, wurde mittels TaqMan™-Low-Density Array mit vorheriger doppelten Prä-amplifizierung, aufgrund der sehr geringen Menge an RNA Menge, quantifiziert. 13 Transkripte, davon zehn Transkripte aus dem 17β-Estradiol-Metabolismus, wurden nicht in Epithelzellen des Bustdrüsengewebes detektiert, während alle Transkripte in geringen Mengen im Homogenat bestimmt wurden. Daher war die Laser-Mirkodissektions-basierte Methode für die Untersuchung des 17β-Estradiol-Metabolismus in Epithelzellen nicht geeignet. Der Vergleich der Transkriptlevel, welche in Estrogenrezeptor α-Aktivierung, Proliferation und Differenzierung eingebunden sind, in Epithelzellen und Homogenat mittels Cluster-Analyse ergab unterschiedliche Clusterzusammensetzungen: Die Cluster des Homogenats unterschieden zwischen den Lobulustypen (alters-bedingte Veränderung der Brustmorphologie). Im Gegensatz dazu wurde in der Clusterzusammensetzung der Epithelzellen zwischen Estrogenrezeptor α-Aktivierung und einer hohen/niedrigen Proliferation und Differenzierung unterschieden. Daher wurden die Daten aus der Homogenatanalyse für weitere Auswertungen des 17β-Estradiol-Metabolismus verwendet, die Untersuchung der Epithelzellproliferation wurde nicht weiter verfolgt.

Zur Abschätzung des Einflusses der Lebensweise und der Polymorphismen auf die spontane Mutationsfrequenz wurde der Random Mutation Capture-Assay, welcher bereits für das p53-Gens in kultivierten Zellen beschrieben war, auf Brustdrüsengewebe übertragen. Zum ersten Mal wurde die Spontanmutationsfrequenz in zwei Proben normalen, humanen Brustdrüsengewebe bestimmt (3,6x10^{-7} und 7,2x10^{-8} pro Basenpaar). Jedoch war die Methode sehr aufwendig und zeitraubend. Daher wurde sie hingehend der Kopienzahlbestimmung und Mutationsbestimmung der Zielsequenz optimiert: Zwei gelelektrophoretische Auftren-

nungen wurden eingespart und das Reaktionsvolumen der PCR zur Bestimmung der Anzahl der Mutanten wurde auf die Hälfte herabgesetzt. Dennoch war die Aufrechterhaltung der Empfindlichkeit des gesamten Assays extrem schwierig und mühsam. Daher war die optimierte Methode immer noch nicht für den Probenumfang in dieser Studie geeignet.

Stattdessen wurde ein metabolisches Netzwerk, welches den 17β-Estradiol-Metabolismus enthielt, in Zusammenarbeit mit dem Lehrstuhl für Bioinformatik der Universität Würzburg, konstruiert. Damit wurden die DNA-Adduktbildungsflüsse als weiterer Marker für Genotoxizität berechnet. Hierzu wurden 14 der 142 enzymatischen Reaktionen aus öffentlichen Datenbanken übernommen, die übrigen Reaktionen wurden durch eine Literatursuche erstellt. Zusätzlich wurden bereits veröffentlichten Netzwerkreaktionen zur Co-Faktorproduktion und zum Energiemetabolismus (bespielsweise Zitronensäurezyklus und oxidative Phosphorylierung) zu dem metabolischen Netzwerk hinzugefügt.

Die DNA-Adduktbildungsflüsse wurden für jede Probe auf Grundlage der quantifizierten Transkriptlevel des 17β-Estradiol-Metabolismus berechnet.

Die DNA-Adduktbildungsflüsse wurden stark von den Transkriptleveln beeinflusst: DNA-Adduktbildungsflüsse variierten um bis zu 9300% unter den Proben. Im Gegensatz dazu hatten die Polymorphismen nur einen geringen Einfluss: Die höchste Zunahme der DNA-Adduktbildungsflüsse von 82,6% wurde bei einer Probe mit einer mittleren Aktivität der NADPH-Chinon-Oxidoreduktase 1 und geringer Aktivität der Catechol-O-Methyltransferase gegenüber dem entsprechenden DNA-Adduktbildungsfluss ohne Berücksichtigung des Genotyps in der Netzwerkberechnung beobachtet. Somit beeinflussen die basalen Transkriptlevel, die wiederum von dem individuellen Lebensstil beeinflusst werden, die Genotoxizität stärker als die Polymorphismen.

Um nun die Frage zu beantworten, ob die Faktoren, die das Brustkrebsrisiko beeinflussen, wie beispielsweise BMI und Rauchen, auch einen Einfluss auf die berechneten DNA-Adduktbildungsflüsse haben, wurde eine lineare Regressionsanalyse durchgeführt. Es zeigte sich, dass ein hoher BMI und der Genotyp -/- der Glutathion-S-Transferase Theta 1 mit einem höheren DNA-Adduktbildungsfluss korreliert waren. Das Alter war negativ mit den DNA- Adduktbildungsflüssen korreliert. Ferner hatten nullipare Spenderinnen höhere DNA-Adduktbildungsflüsse als Frauen, die bereits Kinder hatten. Somit reflektierten die berechneten DNA-Adduktbildungsflüsse die Ergebnisse epidemiologischer Studien zum Brustkrebsrisiko und ergänzten somit einen mechanistischen Beweis zu den deskriptiven Beobachtungen.

Um die Anzahl der Variablen für eine weitere lineare Regressionsanalyse zu reduzieren, wurde eine Korrelationsanalyse durchgeführt. Diese zeigte aber keine Möglichkeit zur Reduzierung der Variablen. Interessanterweise wurden aber Korrelationen von Faktoren, die mit einem erhöhten Brustkrebsrisiko assoziiert sind (Alter und BMI) mit Transkriptleveln,

die in dieser Arbeit quantifiziert wurden, beobachtet. Daher wurden weitere mögliche Assoziationen von Transkriptleveln mit den diskontinuierlichen Variablen, Lobulustyp und Rauchen, in einem zwei Stichprobenvergleich mit anschließender p-Wert-Adjustierung untersucht. Zum einen wurden bereits bekannte Zusammenhänge für Alter und BMI festgestellt, aber auch Neue: *wingless related MMTV integration site*-4 und 5A, beide beteiligt an der Verzweigung der Milchgänge der Brustdrüse und Pregnan-X-Rezeptor, beteiligt an der Induktion verschiedener 17β-Estradiol-metabolisierender Enzyme, waren mit dem Alter korreliert. Ferner war der BMI positiv mit dem *Proliferation related* Ki-67-Antigen korreliert und negativ mit drei Transkripten des 17β-Estradiol-Metabolismus. Rauchen war signifikant mit der Transkriptmenge der Glutathion-S-Transferase Pi 1, welche an der Detoxifizierung von reaktiven Chinonen beteiligt ist, assoziiert. Somit beeinflussten Alter, BMI und Rauchen, die mit einem erhöhten Brustkrebsrisiko assoziiert sind, auch direkt die Transkriptlevel von Enzymen des 17β-Estradiol-Metabolismus im normalen Brustdrüsengewebe.

Um zu untersuchen, ob der gewebsspezifische 17β-Estradiol-Metabolismus entscheidend für die 17β-Estradiol-Gewebespiegel ist, wurde eine lineare Regressionsanalyse durchgeführt. Da dieses statistische Modell in der Anzahl der Variablen begrenzt ist und die Korrelationsanalyse keine Verringerung der Variablen hervorbrachte, wurde die Anzahl der Variablen mittels Hauptkomponentenanalyse der Transkriptlevel, die direkt an der metabolisierung von 17β-Estradiol beteiligt sind, reduziert. Drei Hauptkomponenten wurden identifiziert und zusammen mit Faktoren, wie beispielsweise BMI und Alter, in die statistischen Berechnungen miteinbezogen. Die Veränderung des Lobulustyps, die der Menopause zeitlich voraus geht, war am stärksten (negativ) mit dem 17β-Estradiol-Gewebespiegel korreliert. Abnehmende 17β-Estradiol-Gewebespiegel in der Menopause wurden bisher nur in Tumorgewebe oder normalem Gewebe in Nachbarschaft zum Tumorgewebe beschrieben. Daher könnten die abnehmenden Gewebespiegel auch mit der nachlassenden Funktion der Ovarien begründet sein. Dieses Ergebnis spricht dann allerdings gegen die Hypothese, dass der gewebespezifische Metabolismus entscheidend für die 17β-Estradiol-Gewebespiegel ist. Jedoch wurde auch eine signifikante Korrelation von einer der Hauptkomponenten der 17β-Estradiol metabolisierenden Transkripten mit dem 17β-Estradiol-Gewebespiegel detektiert, was wiederum darauf hinweist, dass auch der gewebespezifische Metabolismus einen Einfluss auf den 17β-Estradiol-Gewebespiegel in der Brustdrüse hat. Folglich scheinen sowohl der gewebespezifische Metabolismus, als auch die Plasmakonzentration von 17β-Estradiol einen Einfluss auf den Gewebespiegel zu haben. Für eine detailliertere Aussage werden jedoch weitere Proben benötigt, um die Datenreduktion durch die Hauptkomponentenanalyse umgehen zu können.

Zusammengefasst wurde der gewebespezifische 17β-Estradiol-Metabolismus in der weiblichen Brustdrüse beschrieben. Unterschiede in der Proliferation von Epithelzellen wur-

den nur in mittels Laser-Mikrodissektion isolierten Epithelzellen widergespiegelt. Es wurde gezeigt, dass Faktoren, die mit Brustkrebsrisiko assoziiert sind (Alter, BMI und altersbedingte Veränderungen in der Brustdrüsenmorphologie), den 17β-Estradiol-Gewebespiegel in der Brustdrüse beeinflussen. Die 17β-Estradiol-vermittelte Genotoxizität wurde mittels bioinformatischer Berechnung der DNA-Adduktflüsse ausgewertet, welche vornehmlich von den individuellen mRNA-Mustern beeinflusst wurde statt von dem individuellen Genotyp. Die DNA-Adduktflüsse korrelierten mit bekannten Brustkrebsrisiko-Faktoren (Alter, Parität, BMI und Polymorphismus der Glutathion-S-Transferase theta 1).

References

Adjei, A. A. and Weinshilboum, R. M. Catecholestrogen sulfation: possible role in carcinogenesis. *Biochem Biophys Res Commun*, 292(2):402–8, 2002.

Afonina, I., Zivarts, M., Kutyavin, I., Lukhtanov, E., Gamper, H., and Meyer, R. B. Efficient priming of pcr with short oligonucleotides conjugated to a minor groove binder. *Nucleic Acids Res*, 25(13):2657–60, 1997.

Albin, N., Massaad, L., Toussaint, C., Mathieu, M. C., Morizet, J., Parise, O., Gouyette, A., and Chabot, G. G. Main drug-metabolizing enzyme systems in human breast tumors and peritumoral tissues. *Cancer Res*, 53(15):3541–6, 1993.

Anders, C. K., Hsu, D. S., Broadwater, G., Acharya, C. R., Foekens, J. A., Zhang, Y., Wang, Y., Marcom, P. K., Marks, J. R., Febbo, P. G., Nevins, J. R., Potti, A., and Blackwell, K. L. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *J Clin Oncol*, 26(20):3324–30, 2008.

Arumugam, A., Subramani, R., Nandy, S., Lopez, R., Boopalan, T., and Lakshmanaswamy, R. Parity and short-term estradiol treatment utilizes similar cellular mechanisms to confer protection against breast cancer. *Cell Physiol Biochem*, 34(2):491–505, 2014.

Asztalos, S., Gann, P. H., Hayes, M. K., Nonn, L., Beam, C. A., Dai, Y., Wiley, E. L., and Tonetti, D. A. Gene expression patterns in the human breast after pregnancy. *Cancer Prev Res (Phila)*, 3(3):301–11, 2010.

Aust, S., Obrist, P., Klimpfinger, M., Tucek, G., Jager, W., and Thalhammer, T. Altered expression of the hormone- and xenobiotic-metabolizing sulfotransferase enzymes 1a2 and 1c1 in malignant breast tissue. *Int J Oncol*, 26(4):1079–85, 2005.

Bandala, C., Floriano-Sanchez, E., Cardenas-Rodriguez, N., Lopez-Cruz, J., and Lara-Padilla, E. Rna expression of cytochrome p450 in mexican women with breast cancer. *Asian Pac J Cancer Prev*, 13(6):2647–53, 2012.

Banerjee, M., Robbins, D., and Chen, T. Modulation of xenobiotic receptors by steroids. *Molecules*, 18(7):7389–406, 2013.

Betticher, D. C., Thatcher, N., Altermatt, H. J., Hoban, P., Ryder, W. D., and Heighway, J. Alternate splicing produces a novel cyclin d1 transcript. *Oncogene*, 11(5):1005–11, 1995.

Bielas, J. H. and Loeb, L. A. Quantification of random genomic mutations. *Nat Methods*, 2 (4):285–90, 2005.

Blair, I. A. Analysis of estrogens in serum and plasma from postmenopausal women: past present, and future. *Steroids*, 75(4-5):297–306, 2010.

Bocker, W., Hungermann, D., and Decker, T. [anatomy of the breast]. *Pathologe*, 30(1): 6–12, 2009.

Bourdon, J. C., Fernandes, K., Murray-Zmijewski, F., Liu, G., Diot, A., Xirodimas, D. P., Saville, M. K., and Lane, D. P. p53 isoforms can regulate p53 transcriptional activity. *Genes Dev*, 19(18):2122–37, 2005.

Bruhn, C., Brockmoller, J., Kerb, R., Roots, I., and Borchert, H. H. Concordance between enzyme activity and genotype of glutathione s-transferase theta (gstt1). *Biochem Pharmacol*, 56(9):1189–93, 1998.

Buchard, A., Sanchez, J. J., Dalhoff, K., and Morling, N. Multiplex pcr detection of gstm1, gstt1, and gstp1 gene variants: simultaneously detecting gstm1 and gstt1 gene copy number and the allelic status of the gstp1 ile105val genetic variant. *J Mol Diagn*, 9(5):612–7, 2007.

Burhans, W. C. and Heintz, N. H. The cell cycle is a redox cycle: linking phase-specific targets to cell fate. *Free Radic Biol Med*, 47(9):1282–93, 2009.

Cairns, J., Wright, C., Cattan, A. R., Hall, A. G., Cantwell, B. J., Harris, A. L., and Horne, C. H. Immunohistochemical demonstration of glutathione s-transferases in primary human breast carcinomas. *J Pathol*, 166(1):19–25, 1992.

Calaf, G. M. and Roy, D. Human drug metabolism genes in parathion-and estrogen-treated breast cells. *Int J Mol Med*, 20(6):875–81, 2007.

Cavalieri, E., Chakravarti, D., Guttenplan, J., Hart, E., Ingle, J., Jankowiak, R., Muti, P., Rogan, E., Russo, J., Santen, R., and Sutter, T. Catechol estrogen quinones as initiators of breast and other human cancers: implications for biomarkers of susceptibility and cancer prevention. *Biochim Biophys Acta*, 1766(1):63–78, 2006.

Cavalieri, E. L. and Rogan, E. G. Unbalanced metabolism of endogenous estrogens in the etiology and prevention of human cancer. *J Steroid Biochem Mol Biol*, 125(3-5):169–80, 2011.

Cecil, A., Rikanovic, C., Ohlsen, K., Liang, C., Bernhardt, J., Oelschlaeger, T. A., Gulder, T., Bringmann, G., Holzgrabe, U., Unger, M., and Dandekar, T. Modeling antibiotic and cytotoxic effects of the dimeric isoquinoline iq-143 on metabolism and its regulation in staphylococcus aureus, staphylococcus epidermidis and human cells. *Genome Biol*, 12(3): R24, 2011.

Cerne, J. Z., Pohar-Perme, M., Novakovic, S., Frkovic-Grazio, S., Stegel, V., and Gersak, K. Combined effect of cyp1b1, comt, gstp1, and mnsod genotypes and risk of postmenopausal breast cancer. *J Gynecol Oncol*, 22(2):110–9, 2011.

Chen, J., Lipska, B. K., Halim, N., Ma, Q. D., Matsumoto, M., Melhem, S., Kolachana, B. S., Hyde, T. M., Herman, M. M., Apud, J., Egan, M. F., Kleinman, J. E., and Weinberger, D. R. Functional analysis of genetic variation in catechol-o-methyltransferase (comt): effects on mrna, protein, and enzyme activity in postmortem human brain. *Am J Hum Genet*, 75(5):807–21, 2004.

Chouinard, S., Tessier, M., Vernouillet, G., Gauthier, S., Labrie, F., Barbier, O., and Belanger, A. Inactivation of the pure antiestrogen fulvestrant and other synthetic estrogen molecules by udp-glucuronosyltransferase 1a enzymes expressed in breast tissue. *Mol Pharmacol*, 69(3):908–20, 2006.

Clarke, R. B. Ovarian steroids and the human breast: regulation of stem cells and cell proliferation. *Maturitas*, 54(4):327–34, 2006.

Claro da Silva, T., Polli, J. E., and Swaan, P. W. The solute carrier family 10 (slc10): beyond bile acid transport. *Mol Aspects Med*, 34(2-3):252–69, 2013.

Coronado, G. D., Beasley, J., and Livaudais, J. Alcohol consumption and the risk of breast cancer. *Salud Publica Mex*, 53(5):440–7, 2011.

Crujeiras, A. B., Diaz-Lagares, A., Carreira, M. C., Amil, M., and Casanueva, F. F. Oxidative stress associated to dysfunctional adipose tissue: a potential link between obesity, type 2 diabetes mellitus and breast cancer. *Free Radic Res*, 47(4):243–56, 2013.

Dawling, S., Roodi, N., Mernaugh, R. L., Wang, X., and Parl, F. F. Catechol-o-methyltransferase (comt)-mediated metabolism of catechol estrogens: comparison of wild-type and variant comt isoforms. *Cancer Res*, 61(18):6716–22, 2001.

Dawling, S., Roodi, N., and Parl, F. F. Methoxyestrogens exert feedback inhibition on cytochrome p450 1a1 and 1b1. *Cancer Res*, 63(12):3127–32, 2003.

Dawling, S., Hachey, D. L., Roodi, N., and Parl, F. F. In vitro model of mammary es-

trogen metabolism: structural and kinetic differences between catechol estrogens 2- and 4-hydroxyestradiol. *Chem Res Toxicol*, 17(9):1258–64, 2004.

de Kok, J. B., Roelofs, R. W., Giesendorf, B. A., Pennings, J. L., Waas, E. T., Feuth, T., Swinkels, D. W., and Span, P. N. Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Lab Invest*, 85(1):154–9, 2005.

Ding, H. X., Fu, Y. Y., Chen, W. X., and Wang, Z. W. Comt val158met polymorphism and breast cancer risk: evidence from 26 case-control studies. *Breast Cancer Res Treat*, 123 (1):265–270, 2010. 627SP Times Cited:8 Cited References Count:40.

Dong, L. M., Potter, J. D., White, E., Ulrich, C. M., Cardon, L. R., and Peters, U. Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. *JAMA*, 299(20): 2423–36, 2008.

Dumas, N. A., He, D., Frost, A. R., and Falany, C. N. Sulfotransferase 2b1b in human breast: differences in subcellular localization in african american and caucasian women. *J Steroid Biochem Mol Biol*, 111(3-5):171–7, 2008.

Dumbgen, L. and Riedwyl, H. On fences and asymmetry in box-and-whiskers plots. *The American Statistician*, 61(4):356–9, 2007.

Dunn, J. F., Nisula, B. C., and Rodbard, D. Transport of steroid hormones: binding of 21 endogenous steroids to both testosterone-binding globulin and corticosteroid-binding globulin in human plasma. *J Clin Endocrinol Metab*, 53(1):58–68, 1981.

Eigeliene, N., Harkonen, P., and Erkkola, R. Effects of estradiol and medroxyprogesterone acetate on expression of the cell cycle proteins cyclin d1, p21 and p27 in cultured human breast tissues. *Cell Cycle*, 7(1):71–80, 2008a.

Eigeliene, N., Harkonen, P., and Erkkola, R. Effects of estradiol and medroxyprogesterone acetate on expression of the cell cycle proteins cyclin d1, p21 and p27 in cultured human breast tissues. *Cell Cycle*, 7(1):71–80, 2008b.

Eisenbrand, G., Metzler, M., and Hennecke, F. J. *Toxikologie fr Naturwissenschaftler und Mediziner Stoffe, Mechanismen, Prfverfahren*. Wiley-VCH, Weinheim, 3., berarb. und aktualisierte aufl edition, 2005.

Eisenthal, R., Danson, M. J., and Hough, D. W. Catalytic efficiency and kcat/km: a useful comparator? *Trends Biotechnol*, 25(6):247–9, 2007.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. *Regression models, methods and applications*. Springer, Berlin [u.a.], 2013.

Feng, S., Cao, Z., and Wang, X. Role of aryl hydrocarbon receptor in cancer. *Biochim Biophys Acta*, 1836(2):197–210, 2013.

Filaire, E., Dupuis, C., Galvaing, G., Aubreton, S., Laurent, H., Richard, R., and Filaire, M. Lung cancer: what are the links with oxidative stress, physical activity and nutrition. *Lung Cancer*, 82(3):383–9, 2013.

Floriano-Sanchez, E., Rodriguez, N. C., Bandala, C., Coballase-Urrutia, E., and Lopez-Cruz, J. Cyp3a4 expression in breast cancer and its association with risk factors in mexican women. *Asian Pac J Cancer Prev*, 15(8):3805–9, 2014.

Fortner, R. T., Hankinson, S. E., Schairer, C., Xu, X., Ziegler, R. G., and Eliassen, A. H. Association between reproductive factors and urinary estrogens and estrogen metabolites in premenopausal women. *Cancer Epidemiol Biomarkers Prev*, 21(6):959–68, 2012.

Gartel, A. L. and Radhakrishnan, S. K. Lost in transcription: p21 repression, mechanisms, and consequences. *Cancer Res*, 65(10):3980–5, 2005.

Gaudet, M. M., Gapstur, S. M., Sun, J., Diver, W. R., Hannan, L. M., and Thun, M. J.

Active smoking and breast cancer risk: original cohort data and meta-analysis. *J Natl Cancer Inst*, 105(8):515–25, 2013.

Goodman, J. E., Jensen, L. T., He, P., and Yager, J. D. Characterization of human soluble high and low activity catechol-o-methyltransferase catalyzed catechol estrogen methylation. *Pharmacogenetics*, 12(7):517–28, 2002.

Guillemette, C., Belanger, A., and Lepine, J. Metabolic inactivation of estrogens in breast tissue by udp-glucuronosyltransferase enzymes: an overview. *Breast Cancer Res*, 6(6): 246–54, 2004.

Hachey, D. L., Dawling, S., Roodi, N., and Parl, F. F. Sequential action of phase i and ii enzymes cytochrome p450 1b1 and glutathione s-transferase p1 in mammary estrogen metabolism. *Cancer Res*, 63(23):8492–9, 2003.

Hallberg, G., Persson, I., Naessen, T., and Magnusson, C. Effects of pre- and postmenopausal use of exogenous hormones on receptor content in normal human breast tissue: a randomized study. *Gynecol Endocrinol*, 24(8):475–80, 2008.

Hanioka, N., Iwabu, H., Hanafusa, H., Nakada, S., and Narimatsu, S. Expression and inducibility of udp-glucuronosyltransferase 1as in mcf-7 human breast carcinoma cells. *Basic Clin Pharmacol Toxicol*, 110(3):253–8, 2012.

Hayes, J. D. and Pulford, D. J. The glutathione s-transferase supergene family: regulation of gst and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit Rev Biochem Mol Biol*, 30(6):445–600, 1995.

He, X. F., Wei, W., Li, S. X., Su, J., Zhang, Y., Ye, X. H., Liu, Y., and Wang, W. Association between the comt val158met polymorphism and breast cancer risk: a meta-analysis of 30,199 cases and 38,922 controls. *Mol Biol Rep*, 39(6):6811–23, 2012. He, Xiao-Feng Wei, Wu Li, Shao-Xia Su, Jiao Zhang, Ying Ye, Xiang-Hua Liu, Yi Wang, Wei Netherlands Mol Biol Rep. 2012 Jun;39(6):6811-23. doi: 10.1007/s11033-012-1506-2.

Howie, A. F., Forrester, L. M., Glancey, M. J., Schlager, J. J., Powis, G., Beckett, G. J., Hayes, J. D., and Wolf, C. R. Glutathione s-transferase and glutathione peroxidase expression in normal and tumour human tissues. *Carcinogenesis*, 11(3):451–8, 1990.

Huang, C. S., Chern, H. D., Chang, K. J., Cheng, C. W., Hsu, S. M., and Shen, C. Y. Breast cancer risk associated with genotype polymorphism of the estrogen-metabolizing genes cyp17, cyp1a1, and comt: a multigenic study on cancer susceptibility. *Cancer Res*, 59(19):4870–5, 1999.

Huguet, E. L., McMahon, J. A., McMahon, A. P., Bicknell, R., and Harris, A. L. Differential expression of human wnt genes 2, 3, 4, and 7b in human breast cell lines and normal and disease states of human breast tissue. *Cancer Res*, 54(10):2615–21, 1994.

Hutson, S. W., Cowen, P. N., and Bird, C. C. Morphometric studies of age related changes in normal human breast and their significance for evolution of mammary cancer. *J Clin Pathol*, 38(3):281–7, 1985.

Insinga, A., Cicalese, A., and Pelicci, P. G. Dna damage response in adult stem cells. *Blood Cells Mol Dis*, 2014.

Jeyapaul, J. and Jaiswal, A. K. Nrf2 and c-jun regulation of antioxidant response element (are)-mediated expression and induction of gamma-glutamylcysteine synthetase heavy subunit gene. *Biochem Pharmacol*, 59(11):1433–9, 2000.

Jiao, H., Liu, C., Guo, W., Peng, L., Chen, Y., and Martin, F. L. Association of cyp1b1 polymorphisms with breast cancer: A case-control study in the han population in ningxia hui autonomous region, p. r. china. *Biomark Insights*, 5:21–7, 2010.

Key, T. J., Appleby, P. N., Reeves, G. K., Roddam, A. W., Helzlsouer, K. J., Alberg, A. J., Rollison, D. E., Dorgan, J. F., Brinton, L. A., Overvad, K., Kaaks, R., Trichopoulou, A., Clavel-Chapelon, F., Panico, S., Duell, E. J., Peeters, P. H., Rinaldi, S., Fentiman, I. S., Dowsett, M., Manjer, J., Lenner, P., Hallmans, G., Baglietto, L., English, D. R., Giles, G. G., Hopper, J. L., Severi, G., Morris, H. A., Hankinson, S. E., Tworoger, S. S., Koenig, K., Zeleniuch-Jacquotte, A., Arslan, A. A., Toniolo, P., Shore, R. E., Krogh, V., Micheli, A., Berrino, F., Barrett-Connor, E., Laughlin, G. A., Kabuto, M., Akiba, S., Stevens, R. G., Neriishi, K., Land, C. E., Cauley, J. A., Lui, L. Y., Cummings, S. R., Gunter, M. J., Rohan, T. E., and Strickler, H. D. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. *Br J Cancer*, 105(5):709–22, 2011.

Khan, S. A., Chatterton, R. T., Michel, N., Bryk, M., Lee, O., Ivancic, D., Heinz, R., Zalles, C. M., Helenowski, I. B., Jovanovic, B. D., Franke, A. A., Bosland, M. C., Wang, J., Hansen, N. M., Bethke, K. P., Dew, A., Coomes, M., and Bergan, R. C. Soy isoflavone supplementation for breast cancer risk reduction: a randomized phase ii trial. *Cancer Prev Res (Phila)*, 5(2):309–19, 2012.

Kim, Y. C., Clark, R. J., Pelegri, F., and Alexander, C. M. Wnt4 is not sufficient to induce lobuloalveolar mammary development. *BMC Dev Biol*, 9:55, 2009.

Kindla, J., Muller, F., Mieth, M., Fromm, M. F., and Konig, J. Influence of non-steroidal anti-inflammatory drugs on organic anion transporting polypeptide (oatp) 1b1- and oatp1b3-mediated drug transport. *Drug Metab Dispos*, 39(6):1047–53, 2011.

Kleivi, K., Teixeira, M. R., Eknaes, M., Diep, C. B., Jakobsen, K. S., Hamelin, R., and Lothe, R. A. Genome signatures of colon carcinoma cell lines. *Cancer Genet Cytogenet*, 155(2):119–31, 2004.

Knupfer, H., Schmidt, R., Stanitz, D., Brauckhoff, M., Schonfelder, M., and Preiss, R. Cyp2c and il-6 expression in breast cancer. *Breast*, 13(1):28–34, 2004.

Kodama, S. and Negishi, M. Sulfotransferase genes: regulation by nuclear receptors in response to xeno/endo-biotics. *Drug Metab Rev*, 45(4):441–9, 2013.

Koepsell, H. The slc22 family with transporters of organic cations, anions and zwitterions. *Mol Aspects Med*, 34(2-3):413–35, 2013.

Kohle, C. and Bock, K. W. Coordinate regulation of human drug-metabolizing enzymes, and conjugate transporters by the ah receptor, pregnane x receptor and constitutive androstane receptor. *Biochem Pharmacol*, 77(4):689–99, 2009.

Kuiper, G. G., Carlsson, B., Grandien, K., Enmark, E., Haggblad, J., Nilsson, S., and Gustafsson, J. A. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology*, 138(3):863–70, 1997.

Kumar, S., Banks, T. W., and Cloutier, S. Snp discovery through next-generation sequencing and its applications. *Int J Plant Genomics*, 2012:831460, 2012.

Kutyavin, I. V., Lukhtanov, E. A., Gamper, H. B., and Meyer, R. B. Oligonucleotides with conjugated dihydropyrroloindole tripeptides: base composition and backbone effects on hybridization. *Nucleic Acids Res*, 25(18):3718–23, 1997.

Labrie, F., Luu-The, V., Labrie, C., and Simard, J. Dhea and its transformation into androgens and estrogens in peripheral target tissues: intracrinology. *Front Neuroendocrinol*, 22(3):185–212, 2001.

Lanigan, F., O'Connor, D., Martin, F., and Gallagher, W. M. Molecular links between mammary gland development and breast cancer. *Cell Mol Life Sci*, 64(24):3159–84, 2007.

Lee, J. M., Anderson, P. C., Padgitt, J. K., Hanson, J. M., Waters, C. M., and Johnson, J. A. Nrf2, not the estrogen receptor, mediates catechol estrogen-induced activation of the antioxidant responsive element. *Biochim Biophys Acta*, 1629(1-3):92–101, 2003.

Lee, M. H. and Lozano, G. Regulation of the p53-mdm2 pathway by 14-3-3 sigma and other proteins. *Semin Cancer Biol*, 16(3):225–34, 2006.

Lehmann, L. and Wagner, J. Gene expression of 17beta-estradiol-metabolizing isozymes: comparison of normal human mammary gland to normal human liver and to cultured human breast adenocarcinoma cells. *Adv Exp Med Biol*, 617:617–24, 2008.

Lehmann, L., Jiang, L., and Wagner, J. Soy isoflavones decrease the catechol-o-methyltransferase-mediated inactivation of 4-hydroxyestradiol in cultured mcf-7 cells. *Carcinogenesis*, 29(2):363–70, 2008.

Lepine, J., Bernard, O., Plante, M., Tetu, B., Pelletier, G., Labrie, F., Belanger, A., and Guillemette, C. Specificity and regioselectivity of the conjugation of estradiol, estrone, and their catecholestrogen and methoxyestrogen metabolites by human uridine diphospho-glucuronosyltransferases expressed in endometrium. *J Clin Endocrinol Metab*, 89(10): 5222–32, 2004.

Li, J., Smyth, P., Cahill, S., Denning, K., Flavin, R., Aherne, S., Pirotta, M., Guenther, S. M., O'Leary, J. J., and Sheils, O. Improved rna quality and taqman pre-amplification method (preamp) to enhance expression analysis from formalin fixed paraffin embedded (ffpe) materials. *BMC Biotechnol*, 8:10, 2008.

Li, Z., Luu-The, V., Poisson-Pare, D., Ouellet, J., Li, S., Labrie, F., and Pelletier, G. Expression of enzymes involved in synthesis and metabolism of estradiol in human breast as studied by immunocytochemistry and in situ hybridization. *Histol Histopathol*, 24(3): 273–82, 2009.

Lonning, P. E., Helle, H., Duong, N. K., Ekse, D., Aas, T., and Geisler, J. Tissue estradiol is selectively elevated in receptor positive breast cancers while tumour estrone is reduced independent of receptor status. *J Steroid Biochem Mol Biol*, 117(1-3):31–41, 2009.

Lu, S. C. Regulation of glutathione synthesis. *Mol Aspects Med*, 2008.

Macias, H. and Hinck, L. Mammary gland development. *Wiley Interdiscip Rev Dev Biol*, 1 (4):533–57, 2012.

MacLaine, N. J. and Hupp, T. R. How phosphorylation controls p53. *Cell Cycle*, 10(6): 916–21, 2011.

Majic, T., Rentzsch, J., Gudlowski, Y., Ehrlich, S., Juckel, G., Sander, T., Lang, U. E., Winterer, G., and Gallinat, J. Comt val108/158met genotype modulates human sensory gating. *Neuroimage*, 55(2):818–24, 2011.

Mao, C., Wang, X. W., Qiu, L. X., Liao, R. Y., Ding, H., and Chen, Q. Lack of association between catechol-o-methyltransferase val108/158met polymorphism and breast cancer risk: a meta-analysis of 25,627 cases and 34,222 controls. *Breast Cancer Res Treat*, 121(3):719–25, 2010a.

Mao, X. Y., Fan, C. F., Zheng, H. C., Wei, J., Yao, F., and Jin, F. p53 nuclear accumulation and eralpha expression in ductal hyperplasia of breast in a cohort of 215 chinese women. *J Exp Clin Cancer Res*, 29:112, 2010b.

McBryan, J., Howlin, J., Napoletano, S., and Martin, F. Amphiregulin: role in mammary gland development and breast cancer. *J Mammary Gland Biol Neoplasia*, 13(2):159–69, 2008.

McNamara, K. M. and Sasano, H. The intracrinology of breast cancer. *J Steroid Biochem*

Mol Biol, 2014.

Meloche, C. A. and Falany, C. N. Expression and characterization of the human 3 beta-hydroxysteroid sulfotransferases (sult2b1a and sult2b1b). *J Steroid Biochem Mol Biol*, 77 (4-5):261–9, 2001.

Mengual, L., Burset, M., Marin-Aguilera, M., Ribal, M. J., and Alcaraz, A. Multiplex preamplification of specific cdna targets prior to gene expression analysis by taqman arrays. *BMC Res Notes*, 1:21, 2008.

Miki, Y., Suzuki, T., Tazawa, C., Blumberg, B., and Sasano, H. Steroid and xenobiotic receptor (sxr), cytochrome p450 3a4 and multidrug resistance gene 1 in human adult and fetal tissues. *Mol Cell Endocrinol*, 231(1-2):75–85, 2005.

Milicevic, Z., Bajic, V., Zivkovic, L., Kasapovic, J., Andjelkovic, U., and Spremo-Potparevic, B. Identification of p53 and its isoforms in human breast carcinoma cells. *ScientificWorld-Journal*, 2014:618698, 2014.

Miller, W. L. and Auchus, R. J. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*, 32(1):81–151, 2011.

Minelli, C., Thompson, J. R., Abrams, K. R., Thakkinstian, A., and Attia, J. How should we use information about hwe in the meta-analyses of genetic association studies? *Int J Epidemiol*, 37(1):136–46, 2008.

Mitrunen, K. and Hirvonen, A. Molecular epidemiology of sporadic breast cancer. the role of polymorphic genes involved in oestrogen biosynthesis and metabolism. *Mutat Res*, 544 (1):9–41, 2003.

Miyoshi, Y. and Noguchi, S. Polymorphisms of estrogen synthesizing and metabolizing genes and breast cancer risk in japanese women. *Biomed Pharmacother*, 57(10):471–81, 2003.

Mobley, J. A., Bhat, A. S., and Brueggemeier, R. W. Measurement of oxidative dna damage by catechol estrogens and analogues in vitro. *Chem Res Toxicol*, 12(3):270–7, 1999.

Modugno, F., Knoll, C., Kanbour-Shakir, A., and Romkes, M. A potential role for the estrogen-metabolizing cytochrome p450 enzymes in human breast carcinogenesis. *Breast Cancer Res Treat*, 82(3):191–7, 2003.

Molnar, F., Kublbeck, J., Jyrkkarinne, J., Prantner, V., and Honkakoski, P. An update on the constitutive androstane receptor (car). *Drug Metabol Drug Interact*, 28(2):79–93, 2013.

Muskhelishvili, L., Thompson, P. A., Kusewitt, D. F., Wang, C., and Kadlubar, F. F. In situ hybridization and immunohistochemical analysis of cytochrome p450 1b1 expression in human normal tissues. *J Histochem Cytochem*, 49(2):229–36, 2001.

Nakajima, T., Shimooka, H., Weixa, P., Segawa, A., Motegi, A., Jian, Z., Masuda, N., Ide, M., Sano, T., Oyama, T., Tsukagoshi, H., Hamanaka, K., and Maeda, M. Immunohistochemical demonstration of 14-3-3 sigma protein in normal human tissues and lung cancers, and the preponderance of its strong expression in epithelial cells of squamous cell lineage. *Pathol Int*, 53(6):353–60, 2003. Nakajima, Takashi Shimooka, Hanako Weixa, Peng Segawa, Atsuki Motegi, Atsushi Jian, Zhang Masuda, Norihiko Ide, Munenori Sano, Takaaki Oyama, Tetsunari Tsukagoshi, Hiroe Hamanaka, Kozue Maeda, Masahiro Australia Pathol Int. 2003 Jun;53(6):353-60.

Nebert, D. W., Roe, A. L., Vandale, S. E., Bingham, E., and Oakley, G. G. Nad(p)h:quinone oxidoreductase (nqo1) polymorphism, exposure to benzene, and predisposition to disease: a huge review. *Genet Med*, 4(2):62–70, 2002.

Nishida, M. The ishikawa cells from birth to the present. *Hum Cell*, 15(3):104–17, 2002.

Niture, S. K., Kaspar, J. W., Shen, J., and Jaiswal, A. K. Nrf2 signaling and cell survival. *Toxicol Appl Pharmacol*, 244(1):37–42, 2010.

Noutsias, M., Rohde, M., Block, A., Klippert, K., Lettau, O., Blunert, K., Hummel, M., Kuhl, U., Lehmkuhl, H., Hetzer, R., Rauch, U., Poller, W., Pauschinger, M., Schultheiss, H. P., Volk, H. D., and Kotsch, K. Preamplification techniques for real-time rt-pcr analyses of endomyocardial biopsies. *BMC Mol Biol*, 9:3, 2008.

Obaidat, A., Roth, M., and Hagenbuch, B. The expression and function of organic anion transporting polypeptides in normal tissues and in cancer. *Annu Rev Pharmacol Toxicol*, 52:135–51, 2012.

Oguztuzun, S., Abu-Hijleh, A., Coban, T., Bulbul, D., Kilic, M., and Iscan, M. Gst isoenzymes in matched normal and neoplastic breast tissue. *Neoplasma*, 58(4):304–10, 2011.

Ohno, S. and Nakajin, S. Determination of mrna expression of human udp-glucuronosyltransferases and application for localization in various human tissues by real-time reverse transcriptase-polymerase chain reaction. *Drug Metab Dispos*, 37(1):32–40, 2009.

O'Lone, R., Frith, M. C., Karlsson, E. K., and Hansen, U. Genomic targets of nuclear estrogen receptors. *Mol Endocrinol*, 18(8):1859–75, 2004.

Park, K., Jang, J., Irimia, D., Sturgis, J., Lee, J., Robinson, J. P., Toner, M., and Bashir, R. 'living cantilever arrays' for characterization of mass of single live cells in fluids. *Lab Chip*, 8(7):1034–41, 2008.

Pasqualini, J. R., Chetrite, G., Blacker, C., Feinstein, M. C., Delalonde, L., Talbi, M., and Maloche, C. Concentrations of estrone, estradiol, and estrone sulfate and evaluation of sulfatase and aromatase activities in pre- and postmenopausal breast cancer patients. *J Clin Endocrinol Metab*, 81(4):1460–4, 1996.

Pemble, S., Schroeder, K. R., Spencer, S. R., Meyer, D. J., Hallier, E., Bolt, H. M., Ketterer, B., and Taylor, J. B. Human glutathione s-transferase theta (gstt1): cdna cloning and the characterization of a genetic polymorphism. *Biochem J*, 300 (Pt 1):271–6, 1994.

Peterson, N. B., Trentham-Dietz, A., Garcia-Closas, M., Newcomb, P. A., Titus-Ernstoff, L., Huang, Y., Chanock, S. J., Haines, J. L., and Egan, K. M. Association of comt haplotypes and breast cancer risk in caucasian women. *Anticancer Res*, 30(1):217–20, 2010.

Pfeiler, G., Treeck, O., Wenzel, G., Goerse, R., Hartmann, A., Schmitz, G., and Ortmann, O. Correlation of body mass index and menopausal status with the intra-tumoral estrogen system in invasive breast cancer. *Gynecol Endocrinol*, 25(3):183–7, 2009.

Pinon, J. D., Labi, V., Egle, A., and Villunger, A. Bim and bmf in tissue homeostasis and malignant disease. *Oncogene*, 27 Suppl 1:S41–52, 2008.

Pirone, J. R., D'Arcy, M., Stewart, D. A., Hines, W. C., Johnson, M., Gould, M. N., Yaswen, P., Jerry, D. J., Smith Schneider, S., and Troester, M. A. Age-associated gene expression in normal breast tissue mirrors qualitative age-at-incidence patterns for breast cancer. *Cancer Epidemiol Biomarkers Prev*, 21(10):1735–44, 2012.

Prindle, M. J., Fox, E. J., and Loeb, L. A. The mutator phenotype in cancer: molecular mechanisms and targeting strategies. *Curr Drug Targets*, 11(10):1296–303, 2010.

Qiao, E., Ji, M., Wu, J., Ma, R., Zhang, X., He, Y., Zha, Q., Song, X., Zhu, L. W., and Tang, J. Expression of the pxr gene in various types of cancer and drug resistance. *Oncol Lett*, 5(4):1093–1100, 2013.

Rajkumar, L., Guzman, R. C., Yang, J., Thordarson, G., Talamantes, F., and Nandi, S. Short-term exposure to pregnancy levels of estrogen prevents mammary carcinogenesis.

Proc Natl Acad Sci U S A, 98(20):11755–9, 2001.

Reeves, G. K., Pirie, K., Green, J., Bull, D., and Beral, V. Reproductive factors and specific histological types of breast cancer: prospective study and meta-analysis. *Br J Cancer*, 100(3):538–44, 2009.

Reuter, M., Kirsch, P., and Hennig, J. Inferring candidate genes for attention deficit hyperactivity disorder (adhd) assessed by the world health organization adult adhd self-report scale (asrs). *J Neural Transm*, 113(7):929–38, 2006.

Rezola, A., Pey, J., Tobalina, L., Rubio, A., Beasley, J. E., and Planes, F. J. Advances in network-based metabolic pathway analysis and gene expression data integration. *Brief Bioinform*, 2014.

Ribieras, S., Tomasetto, C., and Rio, M. C. The ps2/tff1 trefoil factor, from basic research to clinical applications. *Biochim Biophys Acta*, 1378(1):F61–77, 1998.

Richman, P. G. and Meister, A. Regulation of gamma-glutamyl-cysteine synthetase by nonallosteric feedback inhibition by glutathione. *J Biol Chem*, 250(4):1422–6, 1975. 0021-9258 (Print) Journal Article Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S.

RKI and GEKID. Krebs in deutschland 2009/2010. *9. Ausgabe. Robert Koch-Institut (Hrsg) und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (Hrsg).*, 2013.

Rose, D. P. and Vona-Davis, L. Interaction between menopausal status and obesity in affecting breast cancer risk. *Maturitas*, 66(1):33–8, 2010.

Roth, M., Obaidat, A., and Hagenbuch, B. Oatps, oats and octs: the organic anion and cation transporters of the slco and slc22a gene superfamilies. *Br J Pharmacol*, 165(5): 1260–87, 2012.

Rowley, J. W., Oler, A. J., Tolley, N. D., Hunter, B. N., Low, E. N., Nix, D. A., Yost, C. C., Zimmerman, G. A., and Weyrich, A. S. Genome-wide rna-seq analysis of human and mouse platelet transcriptomes. *Blood*, 118(14):e101–11, 2011.

Rudolph, A., Sainz, J., Hein, R., Hoffmeister, M., Frank, B., Forsti, A., Brenner, H., Hemminki, K., and Chang-Claude, J. Modification of menopausal hormone therapy-associated colorectal cancer risk by polymorphisms in sex steroid signaling, metabolism and transport related genes. *Endocr Relat Cancer*, 18(3):371–84, 2011.

Rudolph, A., Hein, R., Hoffmeister, M., Forsti, A., Hemminki, K., Risch, A., Brenner, H., and Chang-Claude, J. Copy number variations of gstt1 and gstm1, colorectal cancer risk and possible effect modification of cigarette smoking and menopausal hormone therapy. *Int J Cancer*, 131(5):E841–8, 2012.

Rushmore, T. H., Morton, M. R., and Pickett, C. B. The antioxidant responsive element. activation by oxidative stress and identification of the dna consensus sequence required for functional activity. *J Biol Chem*, 266(18):11632–9, 1991.

Russo, J. and Russo, I. H. Development of the human breast. *Maturitas*, 49(1):2–15, 2004.

Salvador, J. M., Brown-Clay, J. D., and Fornace, J., A. J. Gadd45 in stress signaling, cell cycle control, and apoptosis. *Adv Exp Med Biol*, 793:1–19, 2013.

Samavat, H. and Kurzer, M. S. Estrogen metabolism and breast cancer. *Cancer Lett*, 2014.

Santen, R. J., Yue, W., and Wang, J. P. Estrogen metabolites and breast cancer. *Steroids*, 2014.

Sazci, A., Ergul, E., Utkan, N. Z., Canturk, N. Z., and Kaya, G. Catechol-o-methyltransferase val 108/158 met polymorphism in premenopausal breast cancer patients. *Toxicology*, 204

(2-3):197–202, 2004.

Schmahl, C., Ludascher, P., Greffrath, W., Kraus, A., Valerius, G., Schulze, T. G., Treutlein, J., Rietschel, M., Smolka, M. N., and Bohus, M. Comt val158met polymorphism and neural pain processing. *PLoS One*, 7(1):e23658, 2012.

Schmelzle, T., Mailleux, A. A., Overholtzer, M., Carroll, J. S., Solimini, N. L., Lightcap, E. S., Veiby, O. P., and Brugge, J. S. Functional role and oncogene-regulated expression of the bh3-only factor bmf in mammary epithelial anoikis and morphogenesis. *Proc Natl Acad Sci U S A*, 104(10):3787–92, 2007.

Schorr, K., Li, M., Krajewski, S., Reed, J. C., and Furth, P. A. Bcl-2 gene family and related proteins in mammary gland involution and breast cancer. *J Mammary Gland Biol Neoplasia*, 4(2):153–64, 1999.

Schwarz, R., Liang, C., Kaleta, C., Kuhnel, M., Hoffmann, E., Kuznetsov, S., Hecker, M., Griffiths, G., Schuster, S., and Dandekar, T. Integrated network reconstruction, visualization and analysis using yanasquare. *BMC Bioinformatics*, 8:313, 2007.

Sergentanis, T. N. and Economopoulos, K. P. Gstt1 and gstp1 polymorphisms and breast cancer risk: a meta-analysis. *Breast Cancer Res Treat*, 121(1):195–202, 2010.

Serra, R., Easter, S. L., Jiang, W., and Baxley, S. E. Wnt5a as an effector of tgfbeta in mammary development and cancer. *J Mammary Gland Biol Neoplasia*, 16(2):157–67, 2011.

Shimada, T., Watanabe, J., Kawajiri, K., Sutter, T. R., Guengerich, F. P., Gillam, E. M., and Inoue, K. Catalytic properties of polymorphic human cytochrome p450 1b1 variants. *Carcinogenesis*, 20(8):1607–13, 1999.

Siegel, D. and Ross, D. Immunodetection of nad(p)h:quinone oxidoreductase 1 (nqo1) in human tissues. *Free Radic Biol Med*, 29(3-4):246–53, 2000.

Siegel, D., McGuinness, S. M., Winski, S. L., and Ross, D. Genotype-phenotype relationships in studies of a polymorphism in nad(p)h:quinone oxidoreductase 1. *Pharmacogenetics*, 9 (1):113–21, 1999.

Siegel, D., Anwar, A., Winski, S. L., Kepa, J. K., Zolman, K. L., and Ross, D. Rapid polyubiquitination and proteasomal degradation of a mutant form of nad(p)h:quinone oxidoreductase 1. *Mol Pharmacol*, 59(2):263–8, 2001.

Simpson, E. R. Sources of estrogen and their importance. *J Steroid Biochem Mol Biol*, 86 (3-5):225–30, 2003.

Singh, S., Chakravarti, D., Edney, J. A., Hollins, R. R., Johnson, P. J., West, W. W., Higginbotham, S. M., Cavalieri, E. L., and Rogan, E. G. Relative imbalances in the expression of estrogen-metabolizing enzymes in the breast tissue of women with breast carcinoma. *Oncol Rep*, 14(4):1091–6, 2005.

Singh, S., Zahid, M., Saeed, M., Gaikwad, N. W., Meza, J. L., Cavalieri, E. L., Rogan, E. G., and Chakravarti, D. Nad(p)h:quinone oxidoreductase 1 arg139trp and pro187ser polymorphisms imbalance estrogen metabolism towards dna adduct formation in human mammary epithelial cells. *J Steroid Biochem Mol Biol*, 117(1-3):56–66, 2009.

Sivik, T., Vikingsson, S., Green, H., and Jansson, A. Expression patterns of 17beta-hydroxysteroid dehydrogenase 14 in human tissues. *Horm Metab Res*, 44(13):949–56, 2012.

Smuc, T. and Rizner, T. L. Expression of 17beta-hydroxysteroid dehydrogenases and other estrogen-metabolizing enzymes in different cancer cell lines. *Chem Biol Interact*, 178(1-3): 228–33, 2009.

Smutny, T., Mani, S., and Pavek, P. Post-translational and post-transcriptional modifications of pregnane x receptor (pxr) in regulation of the cytochrome p450 superfamily. *Curr Drug Metab*, 14(10):1059–69, 2013.

Sobrino, B., Brion, M., and Carracedo, A. Snps in forensic genetics: a review on snp typing methodologies. *Forensic Sci Int*, 154(2-3):181–94, 2005.

Soto, A. M., Sonnenschein, C., Chung, K. L., Fernandez, M. F., Olea, N., and Serrano, F. O. The e-screen assay as a tool to identify estrogens: an update on estrogenic environmental pollutants. *Environ Health Perspect*, 103 Suppl 7:113–22, 1995.

Sprenger, R., Schlagenhaufer, R., Kerb, R., Bruhn, C., Brockmoller, J., Roots, I., and Brinkmann, U. Characterization of the glutathione s-transferase gstt1 deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype-phenotype correlation. *Pharmacogenetics*, 10(6):557–65, 2000.

Starlard-Davenport, A., Lyn-Cook, B., and Radominska-Pandya, A. Identification of udp-glucuronosyltransferase 1a10 in non-malignant and malignant human breast tissues. *Steroids*, 73(6):611–20, 2008.

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., and Stoica, A. 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, 2003. Stoica, Gerald E Franke, Thomas F Moroni, Maria Mueller, Susette Morgan, Elisha Iann, Mary C Winder, Abigail D Reiter, Ronald Wellstein, Anton Martin, Mary Beth Stoica, Adriana P30-CA-51008/CA/NCI NIH HHS/ P50-CA-58185/CA/NCI NIH HHS/ England Oncogene. 2003 Sep 11;22(39):7998-8011.

Stokrova, J., Sloncova, E., Sovova, V., Kucerova, D., Zila, V., Tureckova, J., Vojtechova, M., Korb, J., and Tuhackova, Z. Characterization of four clones derived from human adenocarcinoma cell line, ht29, and analysis of their response to sodium butyrate. *Int J Oncol*, 28(2):559–65, 2006.

Suzuki, T., Miki, Y., Nakata, T., Shiotsu, Y., Akinaga, S., Inoue, K., Ishida, T., Kimura, M., Moriya, T., and Sasano, H. Steroid sulfatase and estrogen sulfotransferase in normal human tissue and breast carcinoma. *J Steroid Biochem Mol Biol*, 86(3-5):449–54, 2003.

Syvanen, A. C., Tilgmann, C., Rinne, J., and Ulmanen, I. Genetic polymorphism of catechol-o-methyltransferase (comt): correlation of genotype with individual variation of s-comt activity and comparison of the allele frequencies in the normal population and parkinsonian patients in finland. *Pharmacogenetics*, 7(1):65–71, 1997.

Tanko, L. B. and Christiansen, C. An update on the antiestrogenic effect of smoking: a literature review with implications for researchers and practitioners. *Menopause*, 11(1): 104–9, 2004.

Terrier, P., Townsend, A. J., Coindre, J. M., Triche, T. J., and Cowan, K. H. An immunohistochemical study of pi class glutathione s-transferase expression in normal human tissue. *Am J Pathol*, 137(4):845–53, 1990.

Thibaudeau, J., Lepine, J., Tojcic, J., Duguay, Y., Pelletier, G., Plante, M., Brisson, J., Tetu, B., Jacob, S., Perusse, L., Belanger, A., and Guillemette, C. Characterization of common ugt1a8, ugt1a9, and ugt2b7 variants with different capacities to inactivate mutagenic 4-hydroxylated metabolites of estradiol and estrone. *Cancer Res*, 66(1):125–33, 2006.

Timofeeva, M., Kropp, S., Sauter, W., Beckmann, L., Rosenberger, A., Illig, T., Jager, B., Mittelstrass, K., Dienemann, H., Bartsch, H., Bickeboller, H., Chang-Claude, J., Risch,

A., and Wichmann, H. E. Genetic polymorphisms of mpo, gstt1, gstm1, gstp1, ephx1 and nqo1 as risk factors of early-onset lung cancer. *Int J Cancer*, 127(7):1547–61, 2010.

Tobacman, J. K., Hinkhouse, M., and Khalkhali-Ellis, Z. Steroid sulfatase activity and expression in mammary myoepithelial cells. *J Steroid Biochem Mol Biol*, 81(1):65–8, 2002.

Traver, R. D., Horikoshi, T., Danenberg, K. D., Stadlbauer, T. H., Danenberg, P. V., Ross, D., and Gibson, N. W. Nad(p)h:quinone oxidoreductase gene expression in human colon carcinoma cells: characterization of a mutation which modulates dt-diaphorase activity and mitomycin sensitivity. *Cancer Res*, 52(4):797–802, 1992.

Traver, R. D., Siegel, D., Beall, H. D., Phillips, R. M., Gibson, N. W., Franklin, W. A., and Ross, D. Characterization of a polymorphism in nad(p)h: quinone oxidoreductase (dt-diaphorase). *Br J Cancer*, 75(1):69–75, 1997.

Tsuchiya, Y., Nakajima, N., Kyo, S., Kanaya, T., Inoue, M., and Yokoi, T. Human cyp1b1 is regulated by estradiol via estrogen receptor. *Cancer Research*, 64(9):3119–3125, 2004.

Vercoutter-Edouart, A. S., Lemoine, J., Le Bourhis, X., Louis, H., Boilly, B., Nurcombe, V., Revillion, F., Peyrat, J. P., and Hondermarck, H. Proteomic analysis reveals that 14-3-3sigma is down-regulated in human breast cancer cells. *Cancer Res*, 61(1):76–80, 2001.

Visvader, J. E. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev*, 23(22):2563–77, 2009.

Visvader, J. E. and Stingl, J. Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev*, 28(11):1143–58, 2014. Visvader, Jane E Stingl, John Genes Dev. 2014 Jun 1;28(11):1143-58. doi: 10.1101/gad.242511.114.

Wagner, J., Jiang, L., and Lehmann, L. Phytoestrogens modulate the expression of 17alpha-estradiol metabolizing enzymes in cultured mcf-7 cells. *Adv Exp Med Biol*, 617:625–32, 2008.

Wang, L., Groves, M. J., Hepburn, M. D., and Bowen, D. T. Glutathione s-transferase enzyme expression in hematopoietic cell lines implies a differential protective role for t1 and a1 isoenzymes in erythroid and for m1 in lymphoid lineages. *Haematologica*, 85(6): 573–9, 2000.

Weisz, J., Fritz-Wolz, G., Gestl, S., Clawson, G. A., Creveling, C. R., Liehr, J. G., and Dabbs, D. Nuclear localization of catechol-o-methyltransferase in neoplastic and nonneoplastic mammary epithelial cells. *Am J Pathol*, 156(6):1841–8, 2000.

Xie, X., Shao, X., Gao, F., Jin, H., Zhou, J., Du, L., Zhang, Y., Ouyang, W., Wang, X., Zhao, L., Zhang, X., and Tang, J. Effect of hyperthermia on invasion ability and tgf-beta1 expression of breast carcinoma mcf-7 cells. *Oncol Rep*, 25(6):1573–9, 2011.

Xu, C., Li, C. Y., and Kong, A. N. Induction of phase i, ii and iii drug metabolism/transport by xenobiotics. *Arch Pharm Res*, 28(3):249–68, 2005.

Yager, J. D. Mechanisms of estrogen carcinogenesis: The role of e2/e1-quinone metabolites suggests new approaches to preventive intervention - a review. *Steroids*, 2014.

Yaghjyan, L. and Colditz, G. A. Estrogens in the breast tissue: a systematic review. *Cancer Causes Control*, 22(4):529–40, 2011.

Yang, H. and Wang, H. Signaling control of the constitutive androstane receptor (car). *Protein Cell*, 5(2):113–23, 2014.

Yang, X., Solomon, S., Fraser, L. R., Trombino, A. F., Liu, D., Sonenshein, G. E., Hestermann, E. V., and Sherr, D. H. Constitutive regulation of cyp1b1 by the aryl hydrocarbon

receptor (ahr) in pre-malignant and malignant mammary tissue. *J Cell Biochem*, 104(2): 402–17, 2008.

Yim, D. S., Parkb, S. K., Yoo, K. Y., Yoon, K. S., Chung, H. H., Kang, H. L., Ahn, S. H., Noh, D. Y., Choe, K. J., Jang, I. J., Shin, S. G., Strickland, P. T., Hirvonen, A., and Kang, D. Relationship between the val158met polymorphism of catechol o-methyl transferase and breast cancer. *Pharmacogenetics*, 11(4):279–86, 2001.

Yu, J., Zhao, Y., Li, B., Sun, L., and Huo, H. 17beta-estradiol regulates the expression of antioxidant enzymes in myocardial cells by increasing nrf2 translocation. *J Biochem Mol Toxicol*, 26(7):264–9, 2012.

Yuan, W., Xu, L., Chen, W., Wang, L., Fu, Z., Pang, D., and Li, D. Evidence on the association between nqo1 pro187ser polymorphism and breast cancer risk in the current studies: a meta-analysis. *Breast Cancer Res Treat*, 125(2):467–72, 2011.

Zhang, B., Beeghly-Fadiel, A., Long, J., and Zheng, W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol*, 12(5):477–88, 2011.

Zhang, J., Schulz, W. A., Li, Y., Wang, R., Zotz, R., Wen, D., Siegel, D., Ross, D., Gabbert, H. E., and Sarbia, M. Association of nad(p)h: quinone oxidoreductase 1 (nqo1) c609t polymorphism with esophageal squamous cell carcinoma in a german caucasian and a northern chinese population. *Carcinogenesis*, 24(5):905–9, 2003.

Zheng, L., Dai, H., Zhou, M., Li, M., Singh, P., Qiu, J., Tsark, W., Huang, Q., Kernstine, K., Zhang, X., Lin, D., and Shen, B. Fen1 mutations result in autoimmunity, chronic inflammation and cancers. *Nat Med*, 13(7):812–9, 2007.

Zhou, C., Verma, S., and Blumberg, B. The steroid and xenobiotic receptor (sxr), beyond xenobiotic metabolism. *Nucl Recept Signal*, 7:e001, 2009.

List of Figures

1 Morphology of the female breast (modified according to Bocker et al., 2009). (A) The intralobular stromal cells are located between the ductules within a lobule. The interlobular stroma is located between lobules. (B) Lobules contain ductules, each lined with glandular epithelial cells (GECs) encircled by myoepithelial cells (MECs). MECs are located at the basal membrane. . . 2

2 Scheme of lobular structure of the human mammary gland. TEB, terminal end bud. LOB, lobule type. 3

3 Simplified scheme of ESR1 activation and proliferation in human mammary gland. NUC, nucleus. ER+, ESR1 containing cells. ER-, cells without ESR1. TGFβR, TGFB receptor. 4

4 Formation of E1 and E2 (modified according to Blair, 2010). DHEA and androstenediol are metabolized to androstenedione and testosterone, respectively, catalyzed by 3β-hydroxysteroid dehydrogenase (HSD), which are then metabolized (CYP19A1, aromatase) to E1 and E2, respectively. E1 and E2 are converted into each other, catalyzed by 17β-HSDs. 6

5 Formation E1 and E2 in premenopausal women. T, transporter. M, metabolites. S, sulfate. Chol, cholesterol. ANDR, androstenedione. SHBG, sex hormone binding globulin or albumin. 8

6 Metabolism of E2 and E1 in normal breast tissue: Activation to genotoxic metabolites (red) and detoxification of genotoxic metabolites (green). Blue arrows indicated polymorphisms known to alter enzyme activity and/or associated with increased breast cancer risk. 12

7 Simplified metabolism of E2 to GSH adducts in normal breast tissue by taking the example of 4-HO-E2. 13

8 Simplified scheme of cellular stress caused by estrogens in human mammary gland. CE, catechol estrogens. QE, estrogen quinones. 16

9 Scheme of RMCA (according to Bielas and Loeb, 2005). 19

10 Decision tree for two-sample comparison. All data is graphically checked on normal distribution and variance homogeneity. 63

11 Photo of human mammary gland tissue sample. 68

12 Representative histological micrograph of a young nulliparous woman with lobule type (LOB) 1, of a parous woman at age of 32 with LOB3 and a parous woman at age of 51 with LOB1$_{parous}$. 70

13 (A) Scheme of the amplified PCR products of RFL PCR for genotyping COMT Val108/158Met before and after restriction digest with *Nla*III. (B) Electrophoretic separation of *Nla*III digested PCR products. 72

14 Scheme and results of allelic discrimination assay for COMT Val108/158Met (rs4680). 74

15 Cluster plot of allelic discrimination for NQO1 Pro187Ser polymorphism of human cell lines. 76

16 (A) Scheme of GSTT1 polymorphism detection. (B) Electrophoretic separation of PCR products. 77

17 Boxplot of mRNA level of COMT, NQO1 and GSTT1 vs. their genotypes of 25 women. 80

18 n0-values (mean±SD) of HPRT of three dependent reactions on one TLDA-WNT32 of 5-50 μl cDNA per slot. 84
19 Frequency distribution of 904 relative SDs (percentage of the mean n0) of n0 (triplicate in one slot of TLDA) of 30 different transcripts dependent on their absolute mean Ct value. 85
20 Frequency distribution of 161 relative SDs (percentage of the mean n0) of n0 (triplicate in one slot of TLDA) of 30 different transcripts dependent on their absolute mean Ct value. 87
21 Experimental design of the investigation on intra and inter TLDA statistical variations. 88
22 Relative SD of the same sample in three slots on one TLDA (intra variation) and two TLDAs (inter variation). 91
23 Experimental design of determination of linearity of preamplification and pre-preamplification. 94
24 Preamplification curves (Ct values of non preamplified cDNA vs. preamplified cDNA) of experiment I with 10, 12 and 14 cycles of preamplification. 96
25 Comparison of transcript levels of NQO1 and GSTP1 in mammary gland homogenate with and without preamplification (PreAmp) of 25 samples. . . 98
26 Relative SD of all transcript levels determined with TLDA-Polymorphism II out of RNA of six different tissue aliquots of breast tissue homogenate (HG) of woman #11 and #30. 103
27 Transcript levels of homogenate (n0/n0 HPRT) with the highest intra sample variation of samples #11 and #30. Presented are the values of the single determination of 23 samples in in comparison to the mean±SD of six aliquots of samples #11 and #30. Transcript levels were determined using TLDA-PolymorphismII and analyzed with SDS software (Chapter 5.2.10). 104
28 Heatmap of levels (n0/n0 HPRT) of transcripts not determined in all samples. 106
29 Comparison of transcript levels between normal human mammary gland tissues and MCF-7 cells cultured with normal culture media or steroid-free culture media. 108
30 Scheme of laser capture microdissection of mammary gland. 115
31 Scheme of cutting-line-setting for laser capture microdissection of GECs+MECs. 116
32 Number of laser dissected elements and area compared to n0 of HPRT transcript of ten tissue sections (points) of nine different tissue aliquots ($\widehat{=}$ nine different donors). 117
33 Transcript levels (n0/n0 HPRT) in isolated MECs+GECS of two tissue sections (I and II) of one sample #29. 118
34 Transcript levels (n0/n0 HPRT) of MECs+GECS (M+G) relative to the respective transcript level of homogenate (HG). 121
35 Cluster dendrograms of the smallest euclidean distance of transcript levels of homogenate and MECs+GECs of the samples. 123
36 Metabolic network calculation of plausibility check. 129
37 Sum of E2 and E1 DNA adduct formation fluxes calculated with transcript levels of 25 mammary gland tissue samples presented as absolute flux values. 131
38 Cluster dendrogram of the smallest euclidean distance of calculated fluxes of DNA adduct formation and transcripts. 133

39 Sum of DNA adduct formation fluxes of E2 and E1 without and with consideration of genotypes. 135
40 Electrophoretic separation of intact DNA and digested DNA. 139
41 (A) Electrophoretical separation of the competitive PCR. (B) Graphical analysis of competitive PCR. 141
42 Scheme of mutant detection procedure. 142
43 Clustered heatmap-representation of Spearman-correlation coefficients. . . . 148
44 Overview of correlation coefficients between transcripts involved in proliferation, differentiation, apoptosis and cell cycle. DSB, ductal side branching. . . 149
45 Transcript levels of NR1I2, WNT4, WNT5A, ESR1, CDKN1A and SULT1E1 in mammary gland homogenate (Chapter 6.3) divided into groups of different ages of the donors. No statistical evaluation was performed, since only the trend of change among the three age groups was of interest. Diamonds, samples. Horizontal line, mean. Vertical lines, SD. 152
46 Representative histological micrograph of sample #5 (assigned to lobule type 1_{parous}). Sections were stained with cresyl violet. Magnification 6.3x. 196
47 Representative histological micrograph of sample #7 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x. 197
48 Representative histological micrograph of sample #10 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x. 197
49 Representative histological micrograph of sample #11 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x. 198
50 Representative histological micrograph of sample #12 (assigned to lobule type 2/3). Sections were stained with cresyl violet. Magnification 6.3x. 198
51 Representative histological micrograph of sample #14 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x. 199
52 Representative histological micrograph of sample #15 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x. 199
53 Representative histological micrograph of sample #24 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x. 200
54 Representative histological micrograph of sample #27 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x. 200
55 Representative histological micrograph of sample #28 (assigned to lobule type 2/3). Sections were stained with H&E. Magnification 6.3x. 201
56 Sequence alignment of the COMT-RFL-PCR uncut product (forward and reverse primer reaction) to the expected sequence (Ensemble genome database.) 203
57 Intra and inter variation of TLDA dependent on method for setting the threshold. 209
58 Box plot of ΔCt values (non preamplified Ct values - preamplified Ct values) of preamplifiaction experiment I. 211
59 Preamplification curves (Ct values of non preamplified cDNA vs. preamplified cDNA) of experiment II with 10, 12 and 14 cycles of preamplification. 212
60 Box plot of ΔCt values (non preamplified Ct values - preamplified Ct values) of preamplifiaction experiment II. 212

61 Pre-preamplification curves with 10, 12 and 14 cycles of pre-preamplification using commercial available RNA transcribed into cDNA and preamplified with the card preamplification pool (experiment I). 213
62 Box plot of ΔCt values (preamplified Ct values - pre-preamplified Ct values) of pre-preamplifiaction experiment I. 213
63 Pre-preamplification curves with 10, 12 and 14 cycles of preamplification using commercial available RNA transcribed into cDNA and preamplified with the card preamplification pool (experiment II). 214
64 Box plot of ΔCt values (preamplified Ct values - pre-preamplified Ct values) of pre-preamplifiaction experiment II. 214
65 Pre-preamplification curves with 10+14, 12+12 and 14+10 cycles of pre-preamplification using commercial available RNA transcribed into cDNA and preamplified with the card preamplification pool (experiment I). 215
66 Box plot of ΔCt values (non preamplified Ct values - pre-preamplified Ct values) of pre-preamplifiaction experiment I. 215
67 Pre-preamplification curves with 10+14, 12+12 and 14+10 cycles of preamplification using commercial available RNA transcribed into cDNA and preamplified with the card preamplification pool (experiment II). 216
68 Box plot of ΔCt values (non preamplified Ct values - pre-preamplified Ct values) of pre-preamplifiaction experiment II. 216
69 Preamplification curves without Ct values of non-preamplifiable transcripts (GSTP1, NQO1, GADD45A, MDM2 and BAD) of experiment I. 217
70 Preamplification curves without Ct values of non-preamplifiable transcripts (GSTP1, NQO1, GADD45A, MDM2 and BAD) of experiment II. 218
71 Comparison of transcript level of enzymes involved in E1/E2 metabolism of homogenate of three aliquots of right and left breast of sample #11 (HG 11) and sample #30 (HG 30). No significant differences (t-test, p value adjustment: Holm). 219
72 Comparison of transcript levels of of three aliquots of right and left breast of sample #11 (HG 11) and sample #30 (HG 30). 220
73 Box plot of preamplification experiment II (12 cycles of preamplification). The nominal ΔCt value was 8.7, whiskers of the box plot were at 7.5 and 9.0 which equals 86% and 104%, respectively. 221
74 Dendrogram (A), cluster classification (B) and characteristics of the two main clusters of transcript levels of homogenate. 228
75 Dendrogram (A), cluster classification (B) and characteristics of the two main clusters of transcript levels of MECs+GECs. 229
76 Cluster analysis of 25 samples with calculated fluxes of CYP1A1, CYP1B1, SULT1A1, SULT1A2, SULT1E1, SULT2A1, COMT, GSTP1, NQO1, HSD17B1, HSD17B2, STS, GCLC and E1/E2 DNA adduct formation fluxes. A: Dendrogramm. B: Cluster classification of the samples. 230
77 Characteristics of all clusters concerning their calculated flux values of enzymes and DNA adducts presented as mean(cluster)-mean(all)/SD(all). . . 231
78 Scree plot of PCA of transcripts of E2 metabolizing enzymes. 235

List of Tables

1 Known substrates among E1/E2 metabolites and precursors of the OATP transporter family (summarized in Roth et al., 2012), OAT transporter (summarized in Koepsell, 2013) and SLC10 transporter (summarized in Claro da Silva et al., 2013) in normal and cancer breast tissue. E1-3-S, E1-3-sulfate. E2-17β-G, E2-17β-glucuronide. Preg-S, pregnenolone sulfate. N, normal breast tissue. C, breast cancer tissue. n.d., not determined. ?, not known. 7

2 Enzymes and isoenzymes metabolizing E1/E2 or metabolites, their substrates in E1/E2 metabolism and whether the enzymes are expressed in normal mammary gland tissue (+) or not (-). If the localization in the mammary gland tissue is known, the "+" was replaced. T, testosterone. MeO-, methoxy-. COMT, catechol-*O*-methyltransferase. ?, no data available. 10

3 Summary of polymorphisms investigated in this work. 22

4 Transcripts with abbreviation and identification number of TaqMan® Gene Expression Assay (assay number) of customized TLDA-WNT-screen designed for the relative quantification of these 16 transcripts including internal card control 18S. 38

5 Transcripts with abbreviation and identification number of TaqMan® Gene Expression Assay (assay number) of customized TLDA-WNT32 designed for the relative quantification of these 32 transcripts including internal card control 18S. 39

6 Transcripts with abbreviation and identification number of TaqMan® Gene Expression Assay (assay number) of customized TLDA designed for the relative quantification of these 52 transcripts, including internal card control 18S. *, only on TLDA-Polymorphisms I. ', only on TLDA-Polymorphism II. °, determined using TaqMan® probe-based real time PCR in 96 well-format. 40

7 Pipetting scheme of reverse transcription using High Capacity Reverse Transcription Kit (Life Technologies). Final volume 20 μl. Template: RNA (Chapter 5.2.6). 47

8 Temperature program for reverse transcription using High Capacity Reverse Transcription Kit (Life Technologies). 47

9 Pipetting scheme of TaqMan® probe-based real time PCR using TaqMan® Gene Expression Assay primers (Life Technologies). Final volume 20 μl. Template: cDNA, preamplified cDNA or pre-preamplified cDNA, if necessary diluted in ncf-water. 48

10 Temperature program for TaqMan® probe-based real time PCR using TaqMan® Gene Expression Assay primers (Life Technologies). 48

11 Pipetting scheme of preamplification (PreAmp, Life Technologies). Final volume 20 μl. Template: cDNA or preamplified cDNA, if necessary diluted in ncf-water. 49

12 Temperature program of preamplification of TaqMan® probe-based real time PCR using preamplification pool (Life Technologies). 49

13 Pipetting scheme of TLDA (Life Technologies). Final volume is 100 μl loaded into one slot. Template (cDNA, preamplified cDNA or pre-preamplified cDNA). 49

14 Temperature program of preamplification of TaqMan® probe-based real time PCR using TLDA-PolymorphismII (Life Technologies). 50
15 Pipetting scheme of allelic discrimination using TaqMan® SNP Genotyping assays (Life Technologies). Final volume 5 μl. Template (DNA). 51
16 Temperature program of the PCR run of the allelic discrimination using TaqMan® SNP Genotyping assays (Life Technologies). 51
17 Pipetting scheme for RFL PCR. Final volume 25 μl. 52
18 Temperature program for RFL PCR. 52
19 Pipetting scheme for RFL PCR. Final volume 25 μl. 53
20 Temperature program of RFL PCR for determination of GSTT1 CNP. . . . 53
21 Pipetting scheme of PCR for synthesis of uracil containing biotinylated probe. Final volume 25 μl. 54
22 Temperature program of PCR for synthesis of biotinylated probe. 54
23 Pipetting scheme for restriction digestion. Final volume 500 μl. Template: human DNA of cells (120 μg) or tissue (300 μg). 55
24 Pipetting scheme for copy number detection using competitive PCR. Final volume 25 μl. Internal standard (competitive DNA) concentrations (copies/μl PCR): 3.2x10^6, 1.6x10^6, 8x10^5, 4x10^5, 2x10^5, 1x10^5, 5x10^4. Dilutions of competitive DNA are performed in ncf-water. 56
25 Temperature program of copy number quantification using competitive PCR. 57
26 Pipetting scheme for dilutions of ct-DNA Stock solution (24 ng/ml) for ct-DNA calibration. conc, concentration. 57
27 I-control software settings for the fluorimetric determination of DNA in the Tecan F200 plate reader (Tecan). 58
28 Primers and TaqMan® probe for real time PCR detection of TP53 copy number. BHQ1, Back Hole Quencher 1® 58
29 Pipetting scheme for copy number detection using TaqMan® probe-based real time PCR. Final volume 20 μl. Template, standard or sample. 58
30 Temperature program of copy number quantification using TaqMan® probe-based real time PCR. 59
31 Pipetting scheme for mutant detection PCR. Final volume 10 μl. 59
32 Temperature program of mutant detection PCR. 60
33 Pipetting scheme for agarose gel electrophoresis. 60
34 Conditions of agarose gel electrophoresis for specific applications. FS, field strength. NEB, New England Biolabs GmbH. 61
35 E2 and E1 levels in human mammary gland homogenates (fmol/g tissue). Determined in working group by means of GC/MS/MS. The sum of E2+E1 and the ratio of E2/E1 are calculated. 62
36 Settings for cofactors and metabolites of E1 and E1 metabolism. 66
37 Distribution of characteristics 30 mammary gland samples. 69
38 Genotypes for COMT Val108/158Met, NQO1 Pro187Ser and GSTT1 of human cell lines. Ishi, Ishikawa cells. 78
39 Polymorphism frequencies (absolute and percent) of COMT Val108/158Met, NQO1 Pro187Ser and GSTT1 CNP of 30 women. 79
40 Non preamplifiable (PA) transcripts in two independent experiments (I and II), Figure 23. 97

41 Non pre-preamplifiable (pPA) transcripts in two independent (I and II) experiments, Figure 23. . . . 100
42 Transcripts with and without preamplification. . . . 101
43 ΔCt values relative to the expected ΔCt value (%Ct_{exp}) of preamplified cDNA for 10, 12 and 14 cycles of preamplification (PA) of HPRT, UGT1A1 and UGT1A10. . . . 101
44 Calculated fluxes of COMT, NQO1, GSTP1, GSTT1, GSTM1, sum of GSTs and formation of DNA adducts for the different genotype scenarios with decreased detoxification. . . . 130
45 Primer chosen for detection of TP53 copy number. . . . 143
46 Summary of the method for copy number determination. . . . 144
47 Distribution of samples analyzed in this study into two categories of each characteristic. LOB, lobule type. n, absolute number of samples. . . . 151
48 Description of the three PCs of transcripts of E1/E2 metabolizing enzymes. . 156
49 Final linear regression models of E2, E1 and E2/E1. . . . 157
50 Final linear regression model of DNA forming adduct flux. . . . 160
51 Characteristics of all samples. . . . 202
52 Genotype of 30 women of COMT Val108/158Met, NQO1 Pro187Ser and GSTT1 CNP. . . . 204
53 Calculation of HWE. Expected and observed frequencies of the COMT Val108/158Met genotypes. . . . 205
54 Calculation of HWE. Expected and observed frequencies of the NQO1 Pro187Ser genotypes. . . . 205
55 Calculation of HWE. Expected and observed frequencies of the GSTT1 CNP genotypes. . . . 205
56 Calculated χ^2 for the three determined Polymorphisms. . . . 206
57 Genotype frequency of COMT Val108/158Met, NQO1 Pro187Ser and GSTT1 CNP. . . . 206
58 Comparison inter and intra TLDA variation. MCF-7 cells, grown in normal culture media, *grown with steroid-free media. . . . 207
59 Ct values of cDNA (20μl per slot) of normal human breast tissue homogenate of sample 0. TLDA from 19.03.2010(File: "2010-03-19 Polymorphismen cDNABrust Homogenat mitPreAmp auswertung.sds"). . . . 210
60 Ct values of smActin, CK18 and HPRT of cDNA of laser dissected glandular epithelial cells of sample #0. . . . 211
61 Concentration of isolated RNA (ng/μl, total volume 65 μl) from human mammary gland tissues. . . . 222
62 Levels of all transcripts investigated in this work relative to HPRT (n0/n0 HPRT). Evaluated with SDS. Samples 0-18. M, mean. . . . 223
63 Levels of all transcripts investigated in this work relative to HPRT (n0/n0 HPRT). Evaluated with SDS. Samples 19-32. M, mean. . . . 223
64 Levels of all transcripts investigated in this work relative to HPRT (n0/n0 HPRT). Evaluated with ExpressionSuite software. Samples 0-20. . . . 225

65 Levels of all transcripts investigated in this work relative to HPRT (n0/n0 HPRT). Evaluated with ExpressionSuite software. Samples 21-32 and mean of MCF-7 cells cultured with normal culture media (DMEM, n=2) or steroid-free media (SF, n=2). 226
66 n0/n0 HPRT of pre-preamplified (14+10 cycles) cDNA MECs+GECs. 227
67 DNA adduct forming fluxes with (+ pol) and without (- pol) consideration of polymorphisms for network calculation. 232
68 Correlation coefficients of correlation analysis of transcripts, age, BMI, E2 and E1 tissue levels, sum of E1 and E2 and ratio E2/E1 with R>0.5. 233
69 Two-sample comparison p values and adjusted p values with method of "Holm" for comparison of transcript levels between smokers and non-smokers (_S) and lobule type 1_{parous} and other lobule types. 234

9 Appendix

9.1 Questionnaire

Version 1 vom 20.09.2012

ÜBERGABEPROTOKOLL (Praxis Prof. Eckert)

Studie: Einfluss Isoflavone auf die Bildung, Aktivierung und Wirkung von 17β-Estradiol in der weiblichen Brustdrüse

Datum: .....

Spenderin (Code): .....

Alter: .....

Anzahl Schwangerschaften: .....

Körpergewicht in kg: .....

Körpergröße: .....

Angaben zum Rauchverhalten:

☐ Raucherin seit: ..... Konsum: .....

☐ Nichtraucherin seit: ..... Konsum bis dato: .....

Angaben zum Alkoholkonsum
(wenn möglich als Menge/Woche und Art der Getränke):

Einnahme von Arzneimitteln mit östrogener Wirkung
(Kontrazeptiva, Hormonsubstitution)? ☐ Ja ☐ Nein

Welches Präparat?

Einnahme von Nahrungsergänzungsmittel
mit östrogener Wirkung? ☐ Ja ☐ Nein

Welche?

Gewebespende wurde von ..... (AK Lehmann) entgegen genommen.

9.2 Lobule types and characterization

Since involution of LOB 1 starts at approx. 40 years, all tissue of parous female donors older than 40 were histological investigated concerning their LOB. Of the 25 samples investigated in this work this included samples #5, #7, #10, #11, #12, #14, #15, #24, #27 and #28. The lobule type of sample #12 and #28 was 2/3 and the remaining samples were lobule type 1_{parous}.

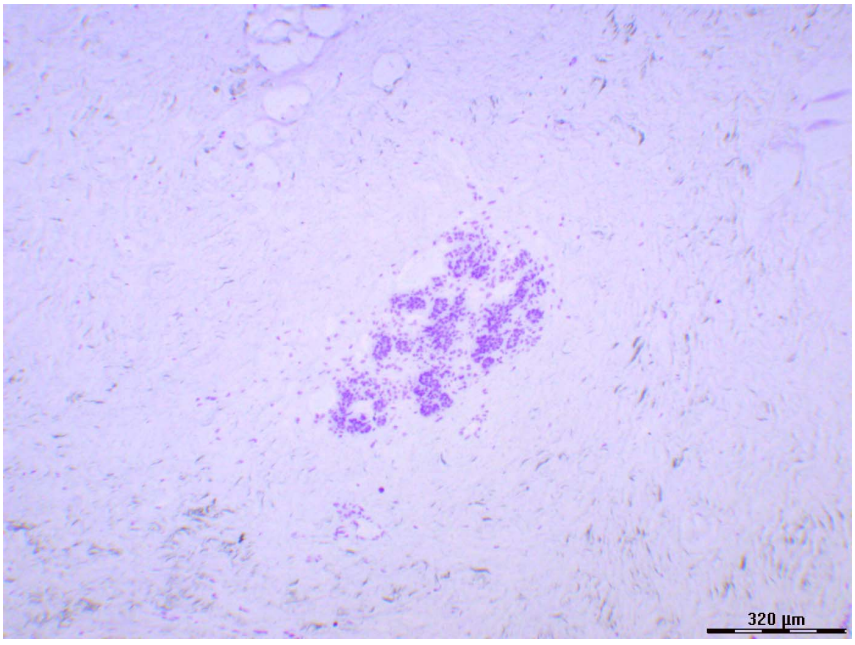


Figure 46: Representative histological micrograph of sample #5 (assigned to lobule type 1_{parous}). Sections were stained with cresyl violet. Magnification 6.3x.

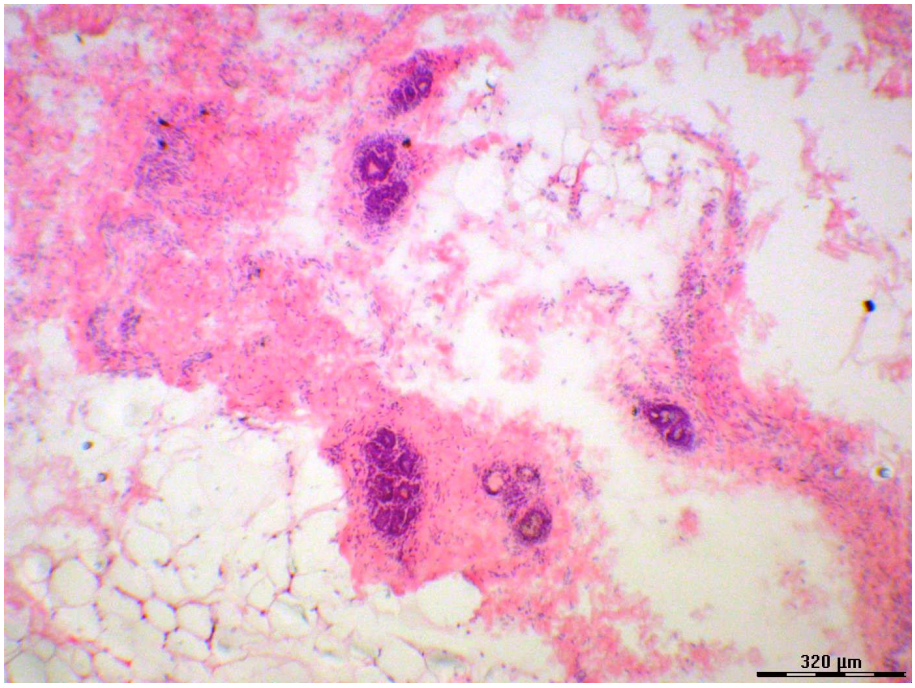


Figure 47: Representative histological micrograph of sample #7 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x.

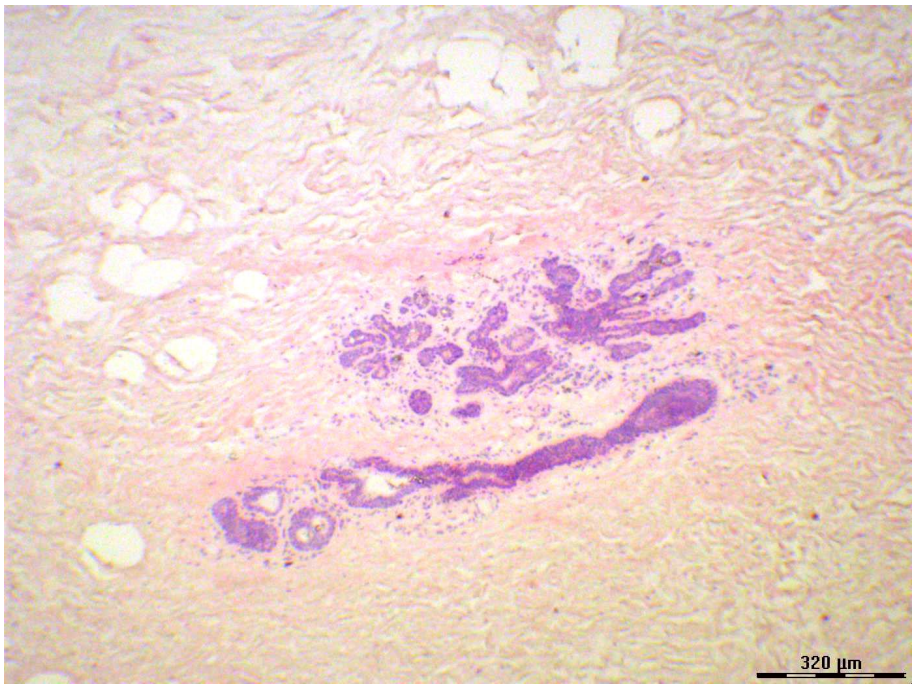


Figure 48: Representative histological micrograph of sample #10 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x.

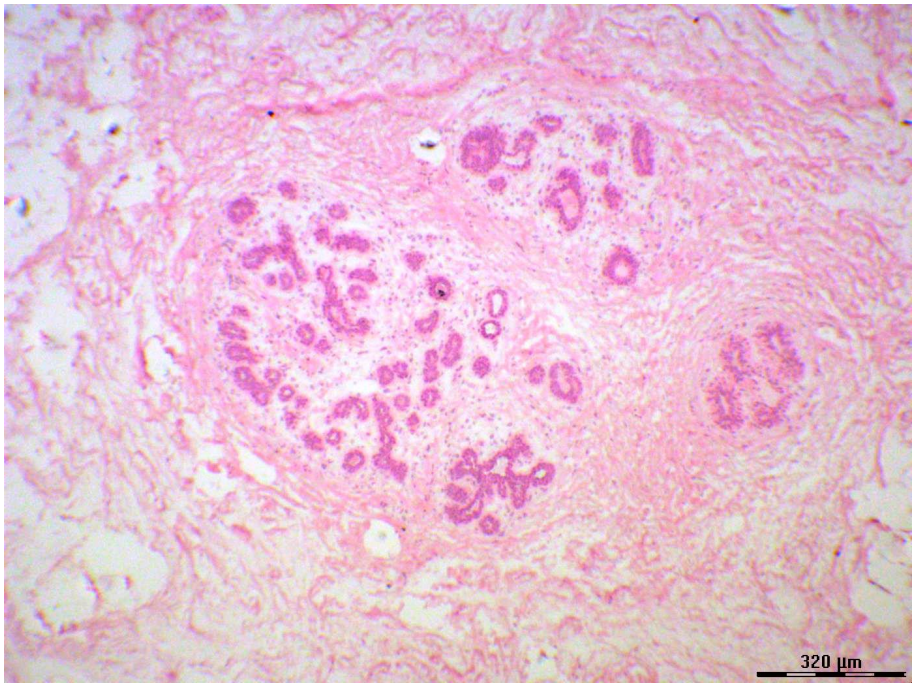


Figure 49: Representative histological micrograph of sample #11 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x.

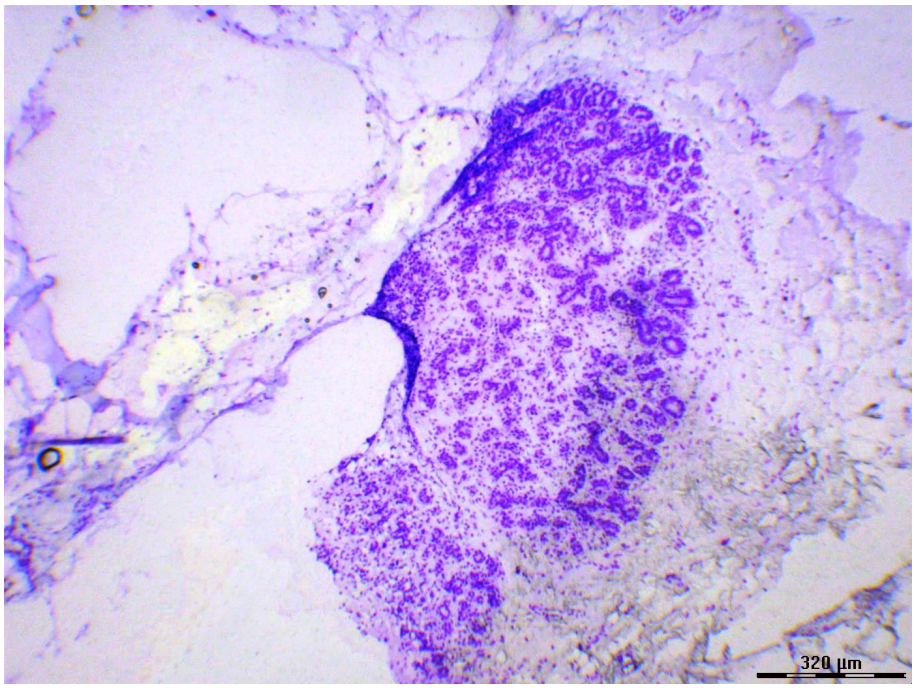


Figure 50: Representative histological micrograph of sample #12 (assigned to lobule type 2/3). Sections were stained with cresyl violet. Magnification 6.3x.

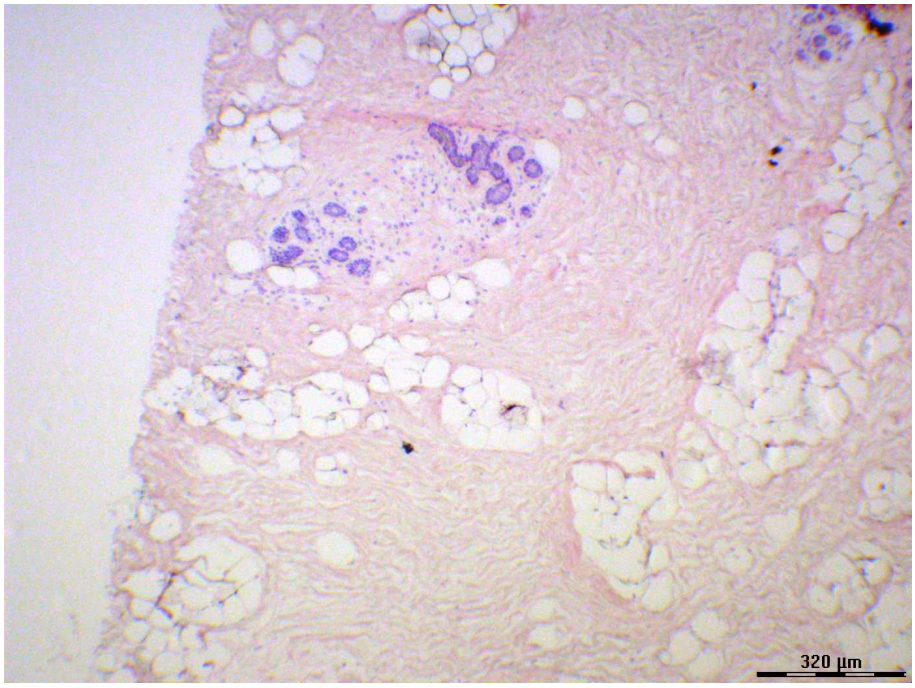


Figure 51: Representative histological micrograph of sample #14 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x.

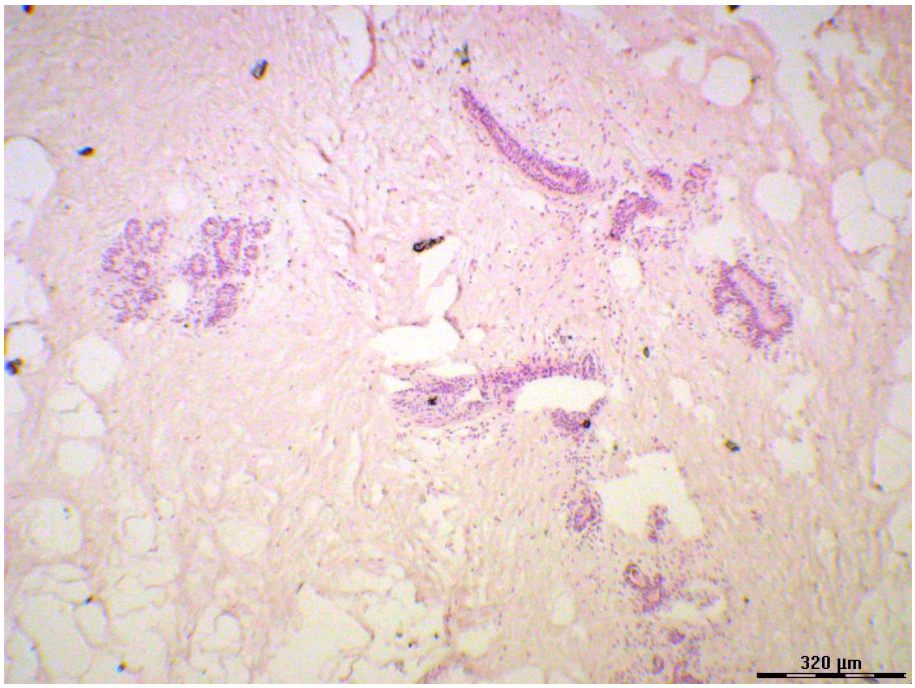


Figure 52: Representative histological micrograph of sample #15 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x.

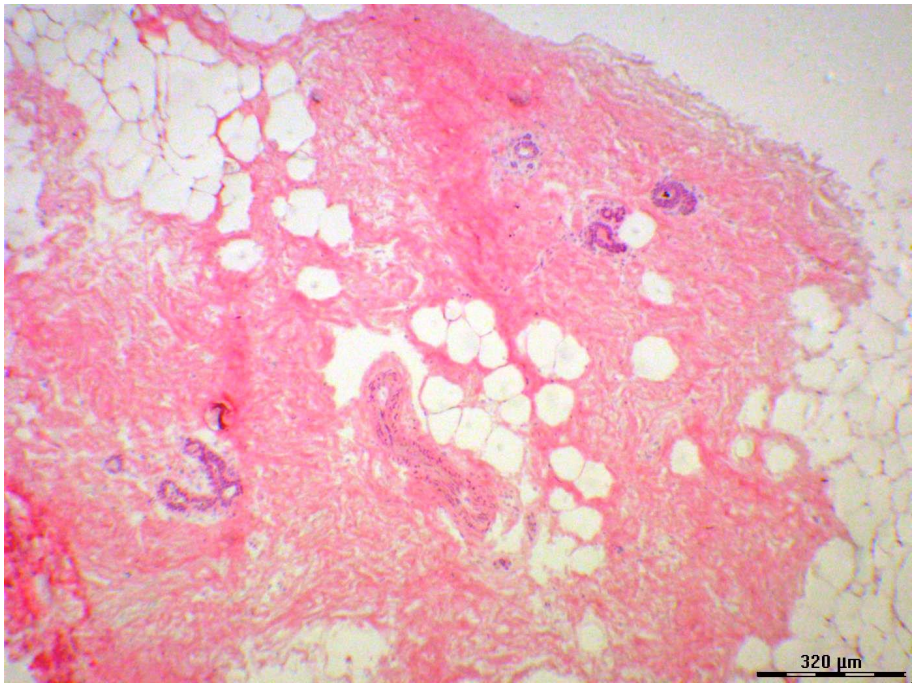


Figure 53: Representative histological micrograph of sample #24 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x.

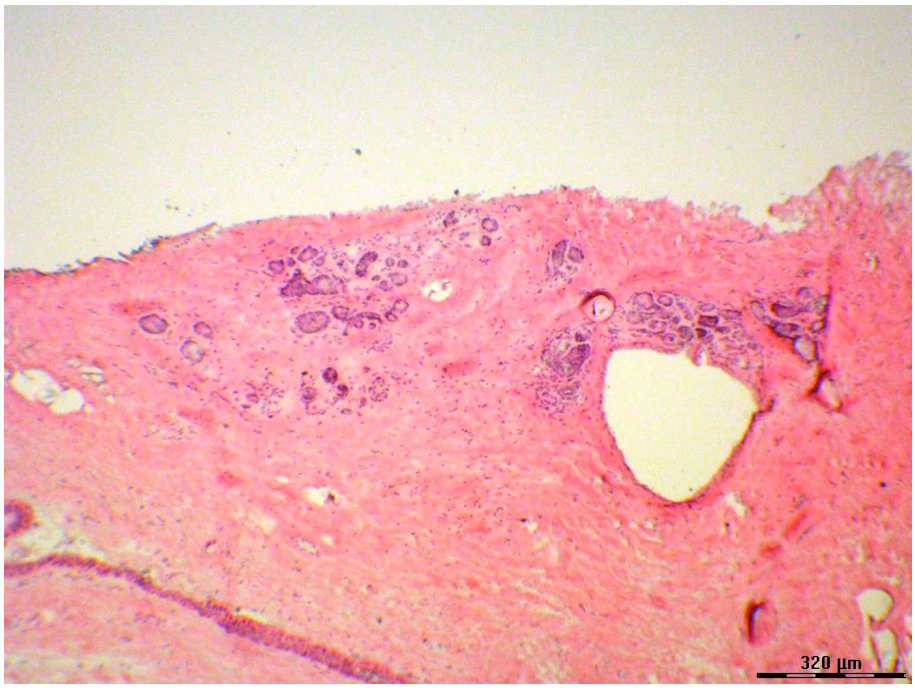


Figure 54: Representative histological micrograph of sample #27 (assigned to lobule type 1_{parous}). Sections were stained with H&E. Magnification 6.3x.

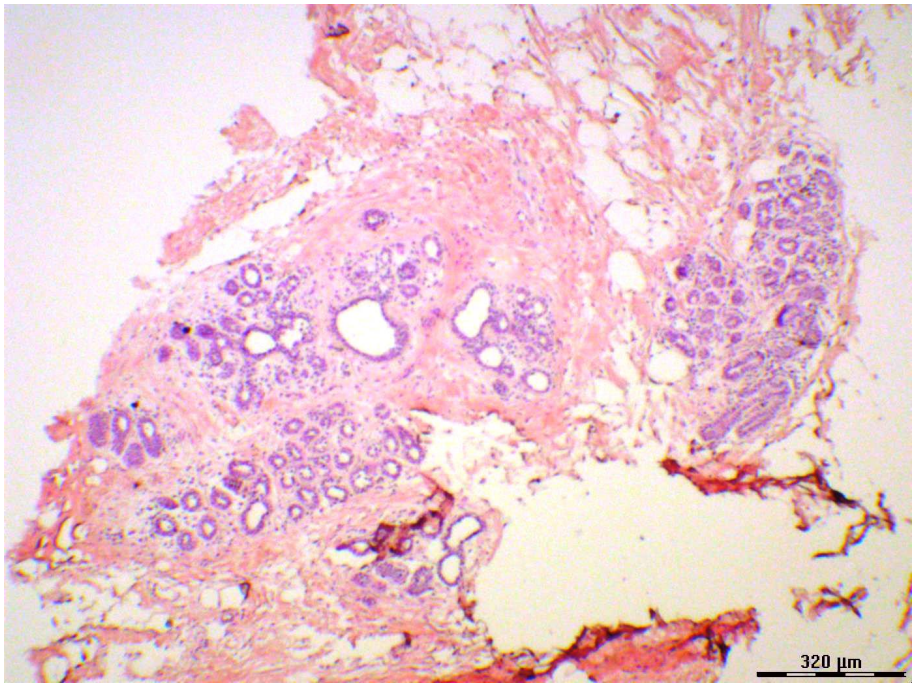


Figure 55: Representative histological micrograph of sample #28 (assigned to lobule type 2/3). Sections were stained with H&E. Magnification 6.3x.

Table 51: Characteristics of all samples. PREG, number of pregnancies. Smoke, smoking habit. HAD, intake of hormone active drugs. DS, dietary supplements. ALC, alcohol consumption.

Sample #	Age	BMI	PREG	Smoke	ALC	HAD	DS	LOB
0	41	22	NA	yes	no	no	no	NA
1	29	27	0	no	no	no	no	1
2	49	25	0	yes	>100 g	yes	no	1
3	30	24	1	yes	no	no	no	2/3
4	18	25	0	no	20-50 g	yes	no	1
5	51	26	3	no	no	no	no	1_{parous}
6	51	28	0	no	20-50 g	no	no	1
7	52	24	1	no	no	no	yes	1_{parous}
8	51	24	0	no	no	no	no	NA
10	51	28	2	no	no	no	no	1_{parous}
11	51	25	2	no	no	no	no	1_{parous}
12	42	27	3	no	20-50 g	yes	no	2/3
13	22	26	1	no	no	no	no	2/3
14	49	26	2	no	no	no	no	1_{parous}
15	53	25	2	no	NA	no	no	1_{parous}
16	27	22	0	yes	20-50 g	no	no	1
17	23	25	0	no	NA	no	no	1
18	35	28	3	no	no	no	no	2/3
19	57	23	NA	no	NA	no	no	NA
20	27	20	0	yes	50-100 g	yes	no	1
21	22	22	0	yes	20-50 g	yes	no	1
24	56	24	2	no	50-100 g	no	no	1_{parous}
25	47	32	0	no	50-100 g	no	no	1
26	44	23	0	yes	20-50 g	yes	no	1
27	54	24	2	no	>100 g	no	no	1_{parous}
28	44	28	2	no	20-50 g	no	no	1
29	32	27	1	no	20-50 g	no	no	2/3
30	30	28	2	no	20-50 g	yes	no	2/3
31	54	24	NA	no	20-50 g	no	no	NA
32	33	27	0	no	no	no	no	1
33	18	26	0	yes	>100 g	yes	no	1

Polymorphisms

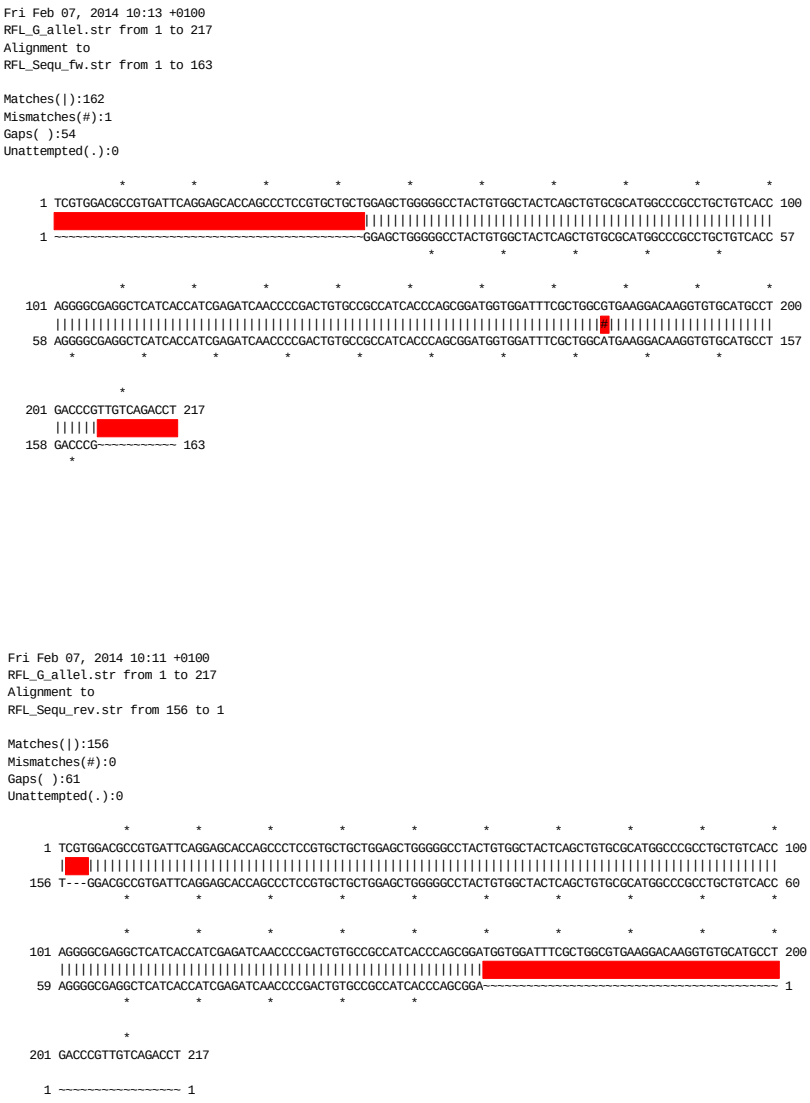


Figure 56: Sequence alignment of the COMT-RFL-PCR uncut product (forward and reverse primer reaction) to the expected sequence (Ensemble genome database.)

Hardy-Weinberg Equilibrium (HWE) was calculated as follows, using the example of COMT Val108/158Met Polymorphism. Hardy-Weinberg equation: $p^2 + 2pq + q^2 = 1$

Frequency of Val/Val genotype = p^2

Frequency of Val/Met genotype = 2pq

Frequency of Met/Met genotype = q^2

Table 52: Genotype of 30 women of COMT Val108/158Met, NQO1 Pro187Ser and GSTT1 CNP.

Breast tissue sample	COMT	GSTT1	NQO1
0	Val/Met	+/+	Pro/Pro
1	Met/Met	+/-	Pro/Pro
2	Val/Val	+/+	Pro/Pro
3	Val/Met	+/+	Pro/Pro
4	Met/Met	+/-	Pro/Pro
5	Met/Met	+/-	Pro/Pro
6	Met/Met	+/+	Pro/Pro
7	Val/Val	+/-	Pro/Pro
8	Val/Val	+/+	Pro/Pro
10	Val/Val	+/-	Pro/Pro
11	Met/Met	+/-	Pro/Pro
12	Val/Val	+/-	Pro/Ser
13	Met/Met	+/-	Pro/Ser
14	Val/Met	+/-	Pro/Ser
15	Met/Met	+/-	Pro/Ser
16	Met/Met	+/+	Pro/Pro
17	Val/Met	-/-	Pro/Ser
18	Met/Met	+/-	Pro/Pro
19	Val/Val	+/-	Pro/Pro
20	Val/Val	-/-	Pro/Pro
21	Met/Met	+/-	Pro/Ser
24	Met/Met	+/+	Pro/Ser
25	Met/Met	+/-	Pro/Pro
26	Met/Met	+/-	Pro/Pro
27	Val/Met	+/+	Pro/Ser
28	Met/Met	+/+	Pro/Pro
29	Val/Met	+/+	Pro/Pro
30	Val/Met	+/-	Pro/Pro
31	Val/Met	+/-	Pro/Pro
32	Val/Met	+/-	Pro/Pro

Using the Hardy-Weinberg equation allele frequencies are calculated:

Frequency of Val = p^2 + 0.5 (2pq) = 0.212 + (0.5 x 0.303) = 0.3635

Frequency of Met = q = 1 - p = 1 - 0.3635 = 0.6365

Using the calculated allele frequencies the expected genotype frequencies are calculated:

Val/Val = p^2 = 0.3635 = 0.132, or 4 individuals in the sample

Val/Met = 2pq = 2 x 0.6365 x 0.3635 = 0.462, or 14 individuals

Met/Met = q^2 = 0.6365 = 0.405, or 13 individuals

COMT NQO1 GSTT1

Table 53: Calculation of HWE. Expected and observed frequencies of the COMT Val108/158Met genotypes.

Genotype	expected	observed
Val/Val	4	7
Val/Met	14	10
Met/Met	13	14

Table 54: Calculation of HWE. Expected and observed frequencies of the NQO1 Pro187Ser genotypes.

Genotype	expected	observed
Pro/Pro	23	22
Pro/Ser	8	9
Ser/Ser	1	0

Table 55: Calculation of HWE. Expected and observed frequencies of the GSTT1 CNP genotypes.

Genotype	expected	observed
+/+	12	10
+/-	14	19
-/-	4	2

Calculation: $\chi^2 = \sum \frac{(observed-expected)^2}{expected}$

The critical value for the chi-square in this case (two degrees of freedom, significance level = 0.05) is 5.991. All calculated chi-square values were less than this critical value. Therefore, the null hypothesis was not rejected and thus all measured polymorphism frequencies were in HWE.

Table 56: Calculated χ^2 for the three determined Polymorphisms.

Polymorphism	χ^2
COMT Val108/158Met	3.5
NQO1 Pro187Ser	1.2
GSTT1 CNP	3.1

Table 57: Genotype frequency of COMT Val108/158Met, NQO1 Pro187Ser and GSTT1 CNP.

Polymorphism	Genotype	Frequency N (%)	Reference
GSTT1 CNP	+/+	576 (32.6%)	(Rudolph et al., 2012)
	+/-	883 (50.0%)	
	-/-	308 (17.4%)	
	+/+ and +/-	113 (80.7%)	(Bruhn et al., 1998)
	-/-	27 (19.3%)	
	+/+	418 (33.5%)	(Timofeeva et al., 2010)
	+/-	617 (49.4%)	
	-/-	214 (17.1%)	
NQO1 Pro187Ser	Pro/Pro	856 (67.6%)	(Timofeeva et al., 2010)
	Pro/Ser and Ser/Ser	411 (32.4%)	
	Pro/Pro	185 (73.4%)	(Zhang et al., 2003)
	Pro/Ser	63 (25.0%)	
	Ser/Ser	4 (1.6%)	
COMT Val108/158Met	Val/Val	171 (25.3%)	(Rudolph et al., 2011)
	Val/Met	343 (50.7%)	
	Met/Met	162 (24.0%)	
	Val/Val	12 (24.0%)	(Schmahl et al., 2012)
	Val/Met	26 (52.0%)	
	Met/Met	12 (24.0%)	
	Val/Val	45 (22.0%)	(Reuter et al., 2006)
	Val/Met	99 (49.0%)	
	Met/Met	59 (29.0%)	
	Val/Val	79 (28.0%)	(Majic et al., 2011)
	Val/Met	155 (54.9%)	
	Met/Met	79 (28.0%)	

9.3 TLDA

Table 58: Comparison inter and intra TLDA variation. MCF-7 cells, grown in normal culture media, *grown with steroid-free media.

	Mean n0/n0 HPRT											
Transcript	cDNA 1	cDNA 2	cDNA3*	cDNA 4*	Mean	SD	Rel. SD (%)	Mean (slot)	Range/2 (slot)	Mean (card)	Range/2 (card)	Comment
CYP1A1	5.50	15.10	8.97	4.49	8.51	4.79	56.24	9.79	5.30	7.24	1.74	
CYP1B1	6.49	5.49	4.75	11.68	7.10	3.13	44.13	8.58	3.09	5.62	0.87	
CYP1A2	n.q.	n.q.	59.08	n.q.	-	-	-	-	-	-	-	ct>34
CYP19A1	n.q.	n.q.	n.q.	n.q.	-	-	-	-	-	-	-	ct>35
UGT1A3/4	n.q.	n.q.	n.q.	n.q.	-	-	-	-	-	-	-	ct>34
UGT2B7	n.d.	n.d.	n.d.	n.d.	-	-	-	-	-	-	-	n.d
UGT1A8	n.d.	n.d.	n.d.	n.d.	-	-	-	-	-	-	-	n.d.
SULT1A1	n.q.	n.q.	n.q.	n.q.	-	-	-	-	-	-	-	ct>34
SULT1A2	15.41	18.75	25.56	42.17	25.47	11.90	46.72	30.46	11.71	20.49	5.07	
SULT1A3/4	7.96	16.05	16.13	12.44	13.14	3.86	29.36	14.24	1.80	12.04	4.08	
18S	mandatory control				-	-	-	-	-	-	-	
SULT1E1	n.q.	n.q.	n.q.	n.q.	-	-	-	-	-	-	-	ct>35
SULT2A1	n.q.	50.35	n.q.	n.q.	-	-	-	-	-	-	-	ct>35
COMT	6.47	5.17	19.18	22.15	13.24	8.67	65.51	13.66	8.49	12.82	6.36	
GSTM1	n.d.	n.d.	n.d.	n.d.	-	-	-	-	-	-	-	n.d.
GSTP1	n.d.	n.d.	n.d.	n.d.	-	-	-	-	-	-	-	n.d.
GSTT1	34.29	7.81	14.80	13.52	17.61	11.53	65.49	10.67	2.85	24.55	9.74	
NQO1	7.62	15.33	8.81	10.41	10.54	3.39	32.16	12.87	2.46	8.21	0.60	
HSD17B1	n.q.	n.q.	32.61	48.63	-	-	-	-	-	-	-	ct>34
HSD17B2	n.d.	n.d.	n.d.	n.d.	-	-	-	-	-	-	-	
STS	10.24	26.46	2.73	14.62	13.51	9.93	73.51	20.54	5.92	6.48	3.75	
ESR1	11.88	1.62	5.80	9.77	7.27	4.53	62.38	5.69	4.07	8.84	3.04	
ESR2	n.q.	n.q.	n.q.	n.q.	-	-	-	-	-	-	-	ct>34
PGR	17.74	12.60	8.89	13.42	13.16	3.63	27.60	13.01	0.41	13.31	4.43	
AHR	7.10	5.01	3.79	5.18	5.27	1.37	25.93	5.10	0.09	5.45	1.65	
NFE2L2	13.61	11.45	3.18	8.06	9.08	4.55	50.12	9.76	1.70	8.39	5.22	
NR1I3	n.q.	n.q.	n.q.	n.q.	-	-	-	-	-	-	-	
NR1I2	n.q.	n.q.	n.q.	n.q.	-	-	-	-	-	-	-	
MKI67	13.25	11.39	14.69	8.65	12.00	2.61	21.73	10.02	1.37	13.97	0.72	
CDKN1B	4.16	6.54	9.34	8.95	7.25	2.40	33.14	7.74	1.21	6.75	2.59	
TFF1	54.23	9.59	7.19	17.49	22.13	21.85	98.76	13.54	3.95	30.71	23.52	
CCND1	5.69	13.16	3.85	22.65	11.34	8.55	75.44	17.91	4.75	4.77	0.92	
GATA3	1.15	1.61	7.13	12.69	5.65	5.42	96.05	7.15	5.54	4.14	2.99	
AREG	6.11	3.62	3.97	10.51	6.05	3.17	52.37	7.07	3.44	5.04	1.07	
TGFB1	42.95	12.01	54.00	12.45	30.35	21.41	70.52	12.23	0.22	48.48	5.53	
WNT5A	39.21	12.13	31.90	4.31	21.89	16.37	74.80	8.22	3.91	35.56	3.66	
WNT4	20.95	15.71	8.17	13.10	14.48	5.33	36.78	14.41	1.31	14.56	6.39	
GCLC	11.38	7.53	14.43	5.14	9.62	4.11	42.74	6.33	1.19	12.91	1.53	
TP53	3.98	6.50	8.41	8.37	6.82	2.09	30.66	7.44	0.93	6.19	2.22	
GADD45A	8.92	13.54	14.58	12.70	12.44	2.47	19.82	13.12	0.42	11.75	2.83	
CDKN1A	21.29	9.87	13.00	13.98	14.53	4.83	33.24	11.92	2.06	17.14	4.14	

	Mean n0/n0 HPRT											
Transcript	cDNA 1	cDNA 2	cDNA3*	cDNA 4*	Mean	SD	Rel. SD (%)	Mean (slot)	Range/2 (slot)	Mean (card)	Range/2 (card)	Comment
SFN	n.d.	n.d.	n.d.	n.d.	-	-	-	-	-	-	-	
MDM2	21.51	16.57	7.60	7.82	13.37	6.85	51.20	12.19	4.38	14.55	6.96	
BAX	9.99	8.63	25.44	30.92	18.75	11.13	59.40	19.78	11.15	17.72	7.72	
BAD	18.95	1.04	11.32	28.33	14.91	11.57	77.62	14.68	13.65	15.13	3.81	
BMF	27.25	10.60	9.50	11.38	14.68	8.41	57.31	10.99	0.39	18.37	8.88	
ARNT	9.28	6.48	5.82	7.30	7.22	1.50	20.77	6.89	0.41	7.55	1.73	
HPRT1	reference gene											

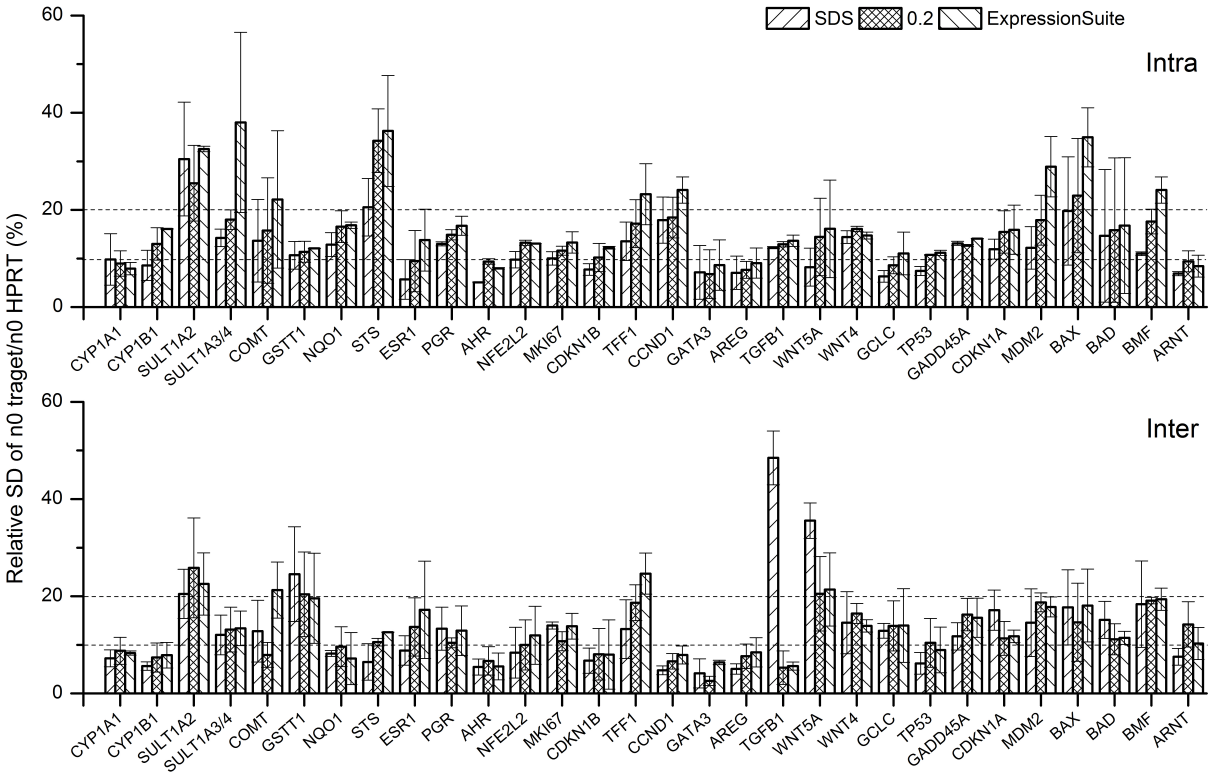


Figure 57: Intra and inter variation of TLDA dependent on method for setting the threshold. Data represent mean±range/2 of two independent samples in duplicate of MCF-7 cells cultured with normal or steroid-free culture media. Out of 47 transcripts 31 transcripts yielded stable Ct values ($<$34 and $>$15. SDS, threshold automatically set for every transcript separate for every TLDA. 0.2, threshold set to "0.2" for every transcript and TLDA. ExpressionSuite, threshold automatically set individual for every transcript to same value for both TLDAs by the ExpressionSuite software (version v1.0.4, Life Technologies).

Table 59: Ct values of cDNA (20μl per slot) of normal human breast tissue homogenate of sample 0. TLDA from 19.03.2010(File: "2010-03-19 Polymorphismen cDNABrust Homogenat mit-PreAmp auswertung.sds").

Gen	Ct value
CYP1A1	Undetermined
CYP1B1	Undetermined
CYP1A2	Undetermined
CYP19A1	37.4
UGT1A4/3	Undetermined
UGT2B7	Undetermined
UGT1A8	Undetermined
SULT1A1	36.3
SULT1A2	31.6
SULT1A3/4	31.1
18S	12.4
SULT1E1	Undetermined
SULT2A1	Undetermined
COMT	28.1
GSTM1	Undetermined
GSTP1	29.5
GSTT1	31.0
NQO1	28.6
HSD17B1	36.3
HSD17B2	31.0
STS	33.1
ESR1	27.3
ESR2	35.0
PGR	29.5
AHR	28.1
NFE2L2	28.7
NR1I3	Undetermined
NR1I2	36.6
MKI67	Undetermined
CDKN1B	27.3
TFF1	27.9
CCND1	30.0
GATA3	26.2
AREG	31.8
TGFB1	29.4
WNT5A	32.0
WNT4	31.7
GCLC	30.0
TP53	28.7
GADD45A	31.5
CDKN1A	29.4
SFN	Undetermined
MDM2	29.2
BAX	29.8
BAD	30.0
BMF	32.5
ARNT	29.4
HPRT	30.6

Table 60: Ct values of smActin, CK18 and HPRT of cDNA of laser dissected glandular epithelial cells of sample #0 (approx. 100 elements containing approx. 500-1000 cells). RNA was isolated (Chapter 5.2.6) and 2.5 μl RNA were reversely transcribed into cDNA (20 μl reaction volume). 1 μl cDNA was amplified in 20 μl real time PCR (File: "02.03.2010 cDNA Dissektion hprt smActin CK18.xlsx").

Target gene	Dilution	ct (replicate1)	ct (replicate2)	ct (replicate3)
smActin	undiluted	n.d.	n.d.	n.d.
	1:2	n.d.	n.d.	n.d.
	1:4	n.d.	n.d.	n.d.
CK18	undiluted	34.7	34.5	n.d.
	1:2	35.6	34.6	35.5
	1:4	n.d.	n.d.	35.6
hprt	undiluted	n.d.	n.d.	n.d.
	1:2	n.d.	n.d.	n.d.
	1:4	n.d.	n.d.	n.d.

9.3.1 Preamplification

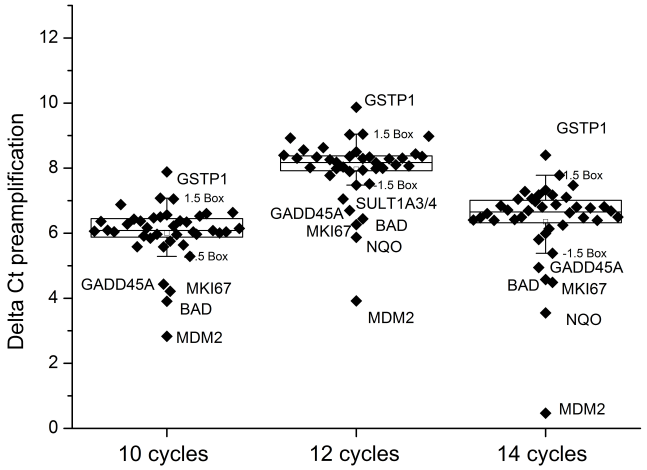


Figure 58: Box plot of ΔCt values (non preamplified Ct values - preamplified Ct values) of preamplifiaction experiment I. Outliers ($\widehat{=}$ points outside of whiskers) were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

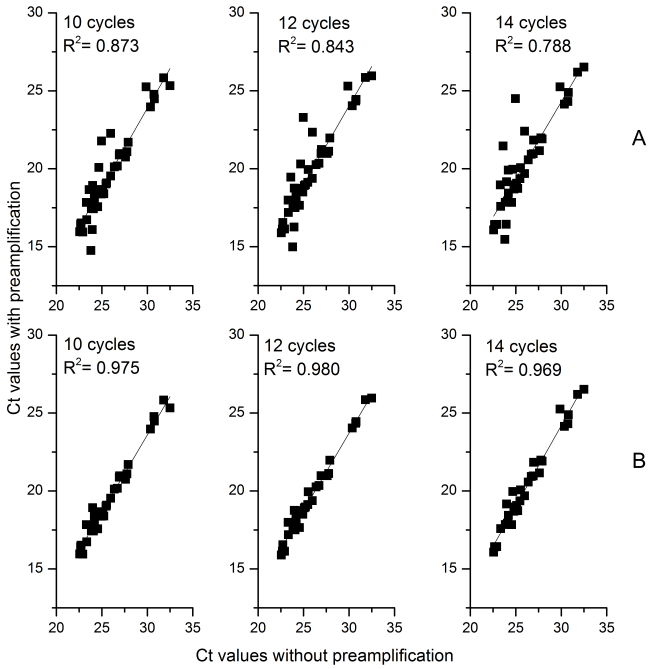


Figure 59: Preamplification curves (Ct values of non preamplified cDNA vs. preamplified cDNA) of experiment II with 10, 12 and 14 cycles of preamplification. Commercial available RNA was reversely transcribed into cDNA and preamplified. Then real time PCR with TLDA-Polymorphisms II with and without preamplification was performed. (A) Ct values of all quantifiable transcripts. (B) Ct values of quantifiable transcripts excluding outliers. Outliers were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

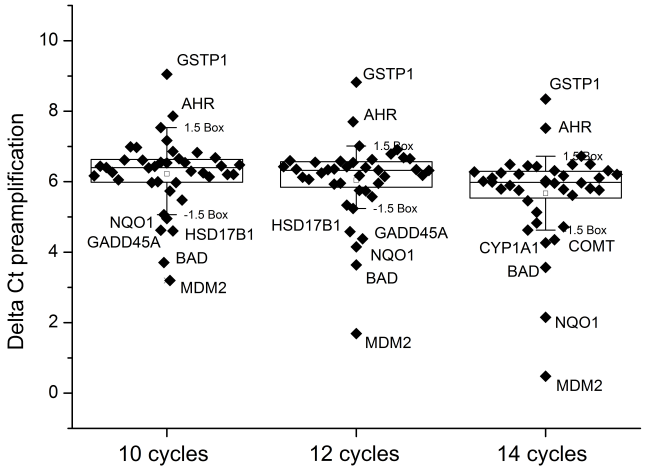


Figure 60: Box plot of ΔCt values (non preamplified Ct values - preamplified Ct values) of preamplifiaction experiment II. Outliers ($\widehat{=}$ points outside of whiskers) were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

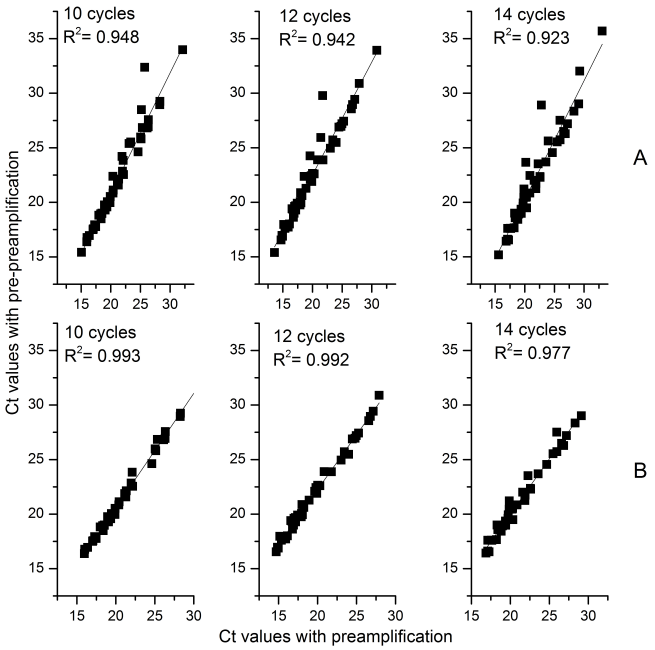


Figure 61: Pre-preamplification curves with 10, 12 and 14 cycles of pre-preamplification using commercial available RNA transcribed into cDNA and preamplified with the card preamplification pool (experiment I). Ct values of pre-preamplification plotted against their corresponding preamplification Ct values (Ct values of 14+10 cycles of pre-preamplification to Ct values of 14 cycles preamplification, 12+12 to 12 and 10+14 to 10). (A) Including all quantifiable transcripts. (B) Excluding transcripts, which are not preamplifiable. Outliers were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

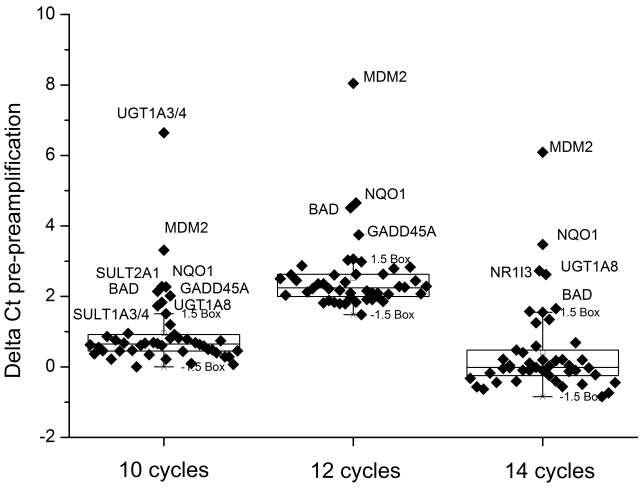


Figure 62: Box plot of ΔCt values (preamplified Ct values - pre-preamplified Ct values) of pre-preamplifiaction experiment I. Outliers ($\widehat{=}$ points outside of whiskers) were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

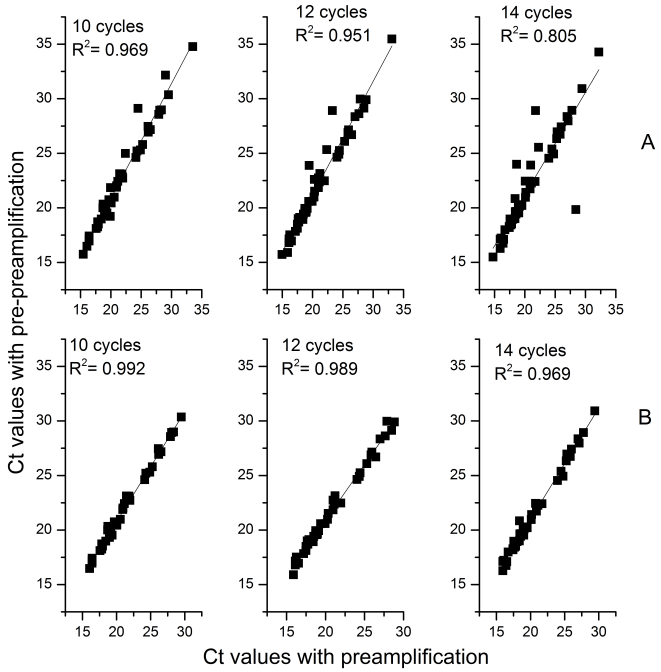


Figure 63: Pre-preamplification curves with 10, 12 and 14 cycles of preamplification using commercial available RNA transcribed into cDNA and preamplified with the card preamplification pool (experiment II). Ct values of pre-preamplification plotted against their corresponding preamplification Ct values (Ct values of 14+10 cycles of pre-preamplification to Ct values of 14 cycles preamplification, 12+12 to 12 and 10+14 to 10). (B) Excluding transcripts, which are not preamplifiable. Outliers were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

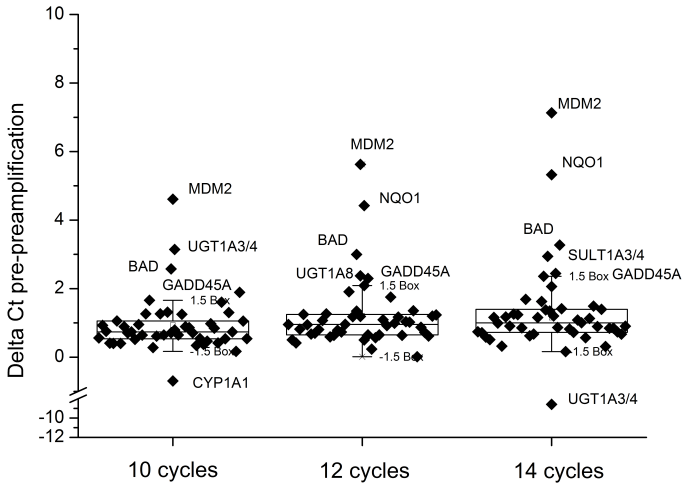


Figure 64: Box plot of ΔCt values (preamplified Ct values - pre-preamplified Ct values) of pre-preamplifiaction experiment II. Outliers ($\widehat{=}$ points outside of whiskers) were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

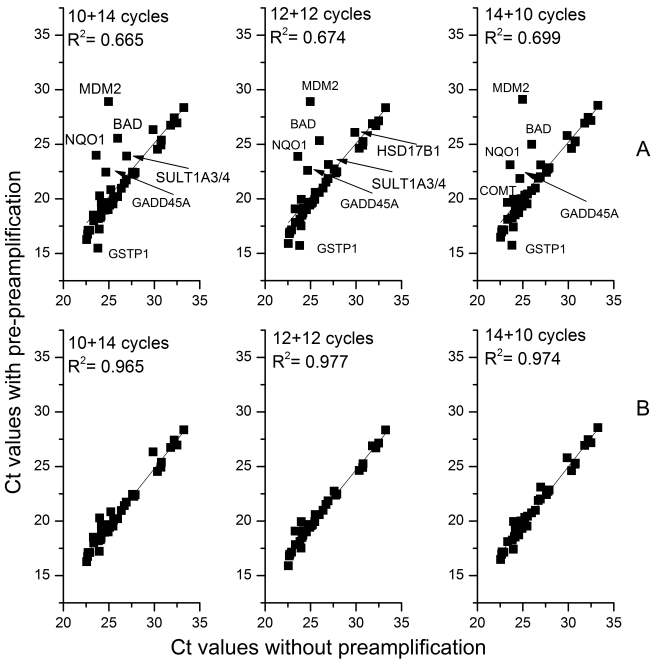


Figure 65: Pre-preamplification curves with 10+14, 12+12 and 14+10 cycles of pre-preamplification using commercial available RNA transcribed into cDNA and preamplified with the card preamplification pool (experiment I). Ct values of pre-preamplification plotted against their corresponding Ct values without preamplification. (A) Including all quantifiable transcripts. (B) Excluding transcripts, which were not preamplifiable. Outliers were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

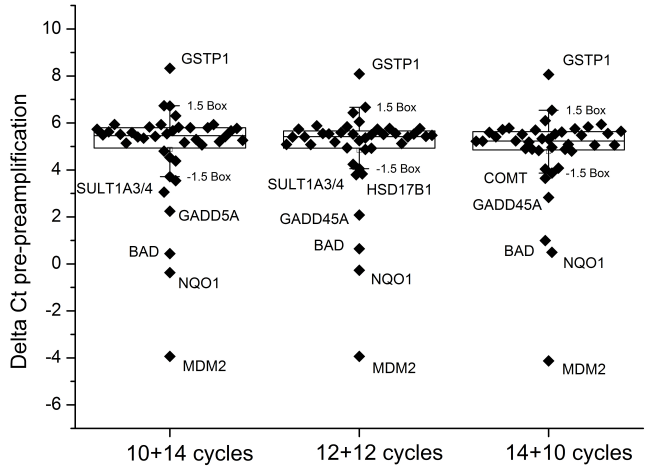


Figure 66: Box plot of ΔCt values (non preamplified Ct values - pre-preamplified Ct values) of pre-preamplifiaction experiment I. Outliers ($\widehat{=}$ points outside of whiskers) were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

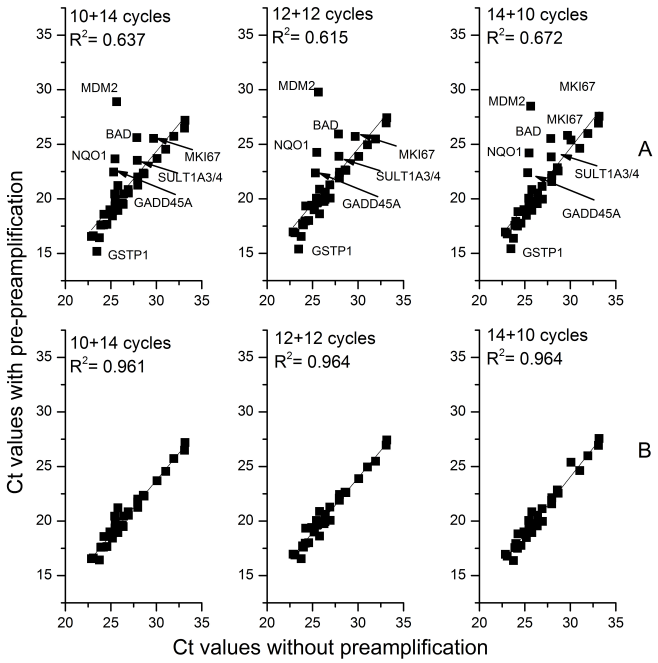


Figure 67: Pre-preamplification curves with 10+14, 12+12 and 14+10 cycles of preamplification using commercial available RNA transcribed into cDNA and preamplified with the card preamplification pool (experiment II). Ct values of pre-preamplification plotted against their corresponding Ct values without preamplification. (A) Including all quantifiable transcripts. (B) Excluding transcripts, which were not preamplifiable. Outliers were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

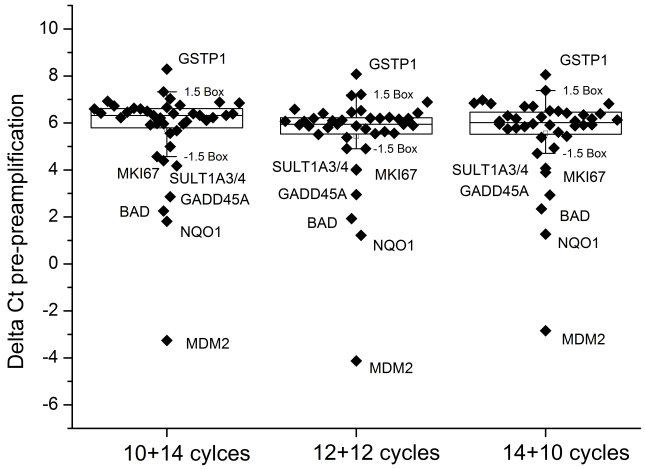


Figure 68: Box plot of ΔCt values (non preamplified Ct values - pre-preamplified Ct values) of pre-preamplifiaction experiment II. Outliers ($\widehat{=}$ points outside of whiskers) were identified using Tukey's rule of "1.5x inter quartile range" (Dumbgen and Riedwyl, 2007).

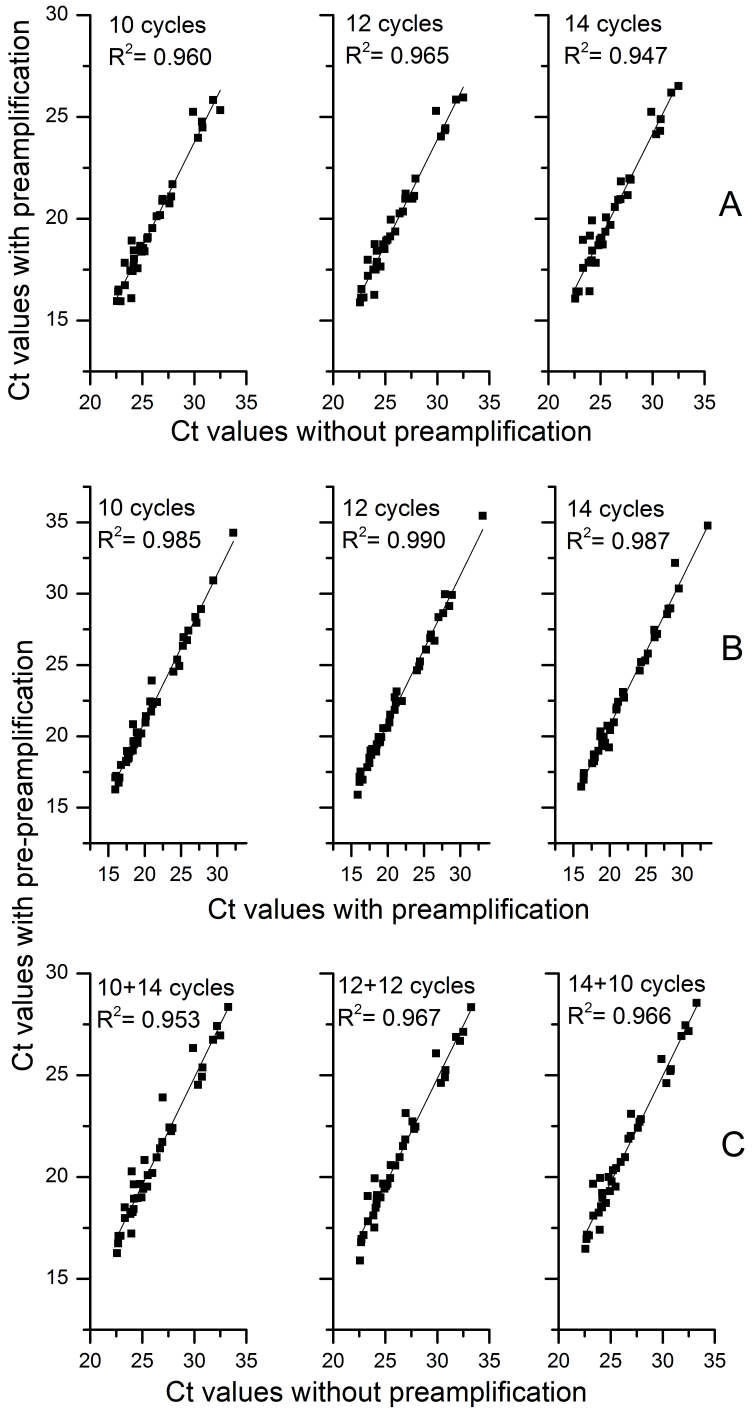


Figure 69: Preamplification curves without Ct values of non-preamplifiable transcripts (GSTP1, NQO1, GADD45A, MDM2 and BAD) using commercial available RNA transcribed into cDNA and preamplified with the card preamplification pool (experiment I). (A) Preamplification curves of 10, 12 and 14 cycles. Ct values of preamplification plotted against their corresponding Ct values without preamplification. (B) Pre-preamplification curves of 10, 12 and 14 cycles. Ct values of pre-preamplification plotted against their corresponding Ct values of preamplification (Ct values of 14+10 cycles of pre-preamplification to Ct values of 14 cycles preamplification, 12+12 to 12 and 10+14 to 10). (C) Pre-preamplification curves of 10+14, 12+12 and 14+10 cycles. Ct values of pre-preamplification plotted against their corresponding Ct values without preamplification.

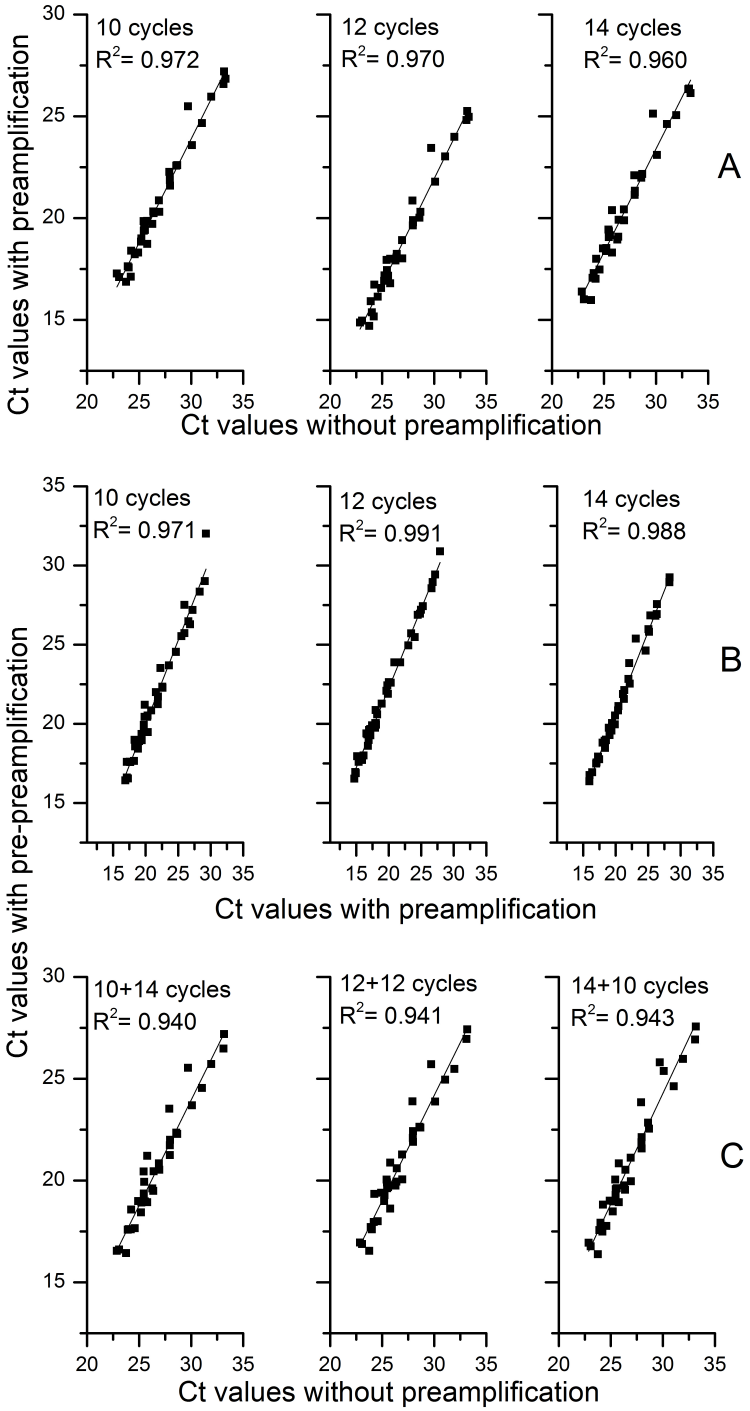


Figure 70: Preamplification curves without Ct values of non-preamplifiable transcripts (GSTP1, NQO1, GADD45A, MDM2 and BAD) using commercial available RNA transcribed into cDNA and preamplified with the card preamplification pool (experiment II). (A) Preamplification curves of 10, 12 and 14 cycles. Ct values of preamplification plotted against their corresponding Ct values without preamplification. (B) Pre-preamplification curves of 10, 12 and 14 cycles. Ct values of pre-preamplification plotted against their corresponding Ct values of preamplification (Ct values of 14+10 cycles of pre-preamplification to Ct values of 14 cycles preamplification, 12+12 to 12 and 10+14 to 10). (C) Pre-preamplification curves of 10+14, 12+12 and 14+10 cycles. Ct values of pre-preamplification plotted against their corresponding Ct values without preamplification.

9.3.2 Sample-taking sites

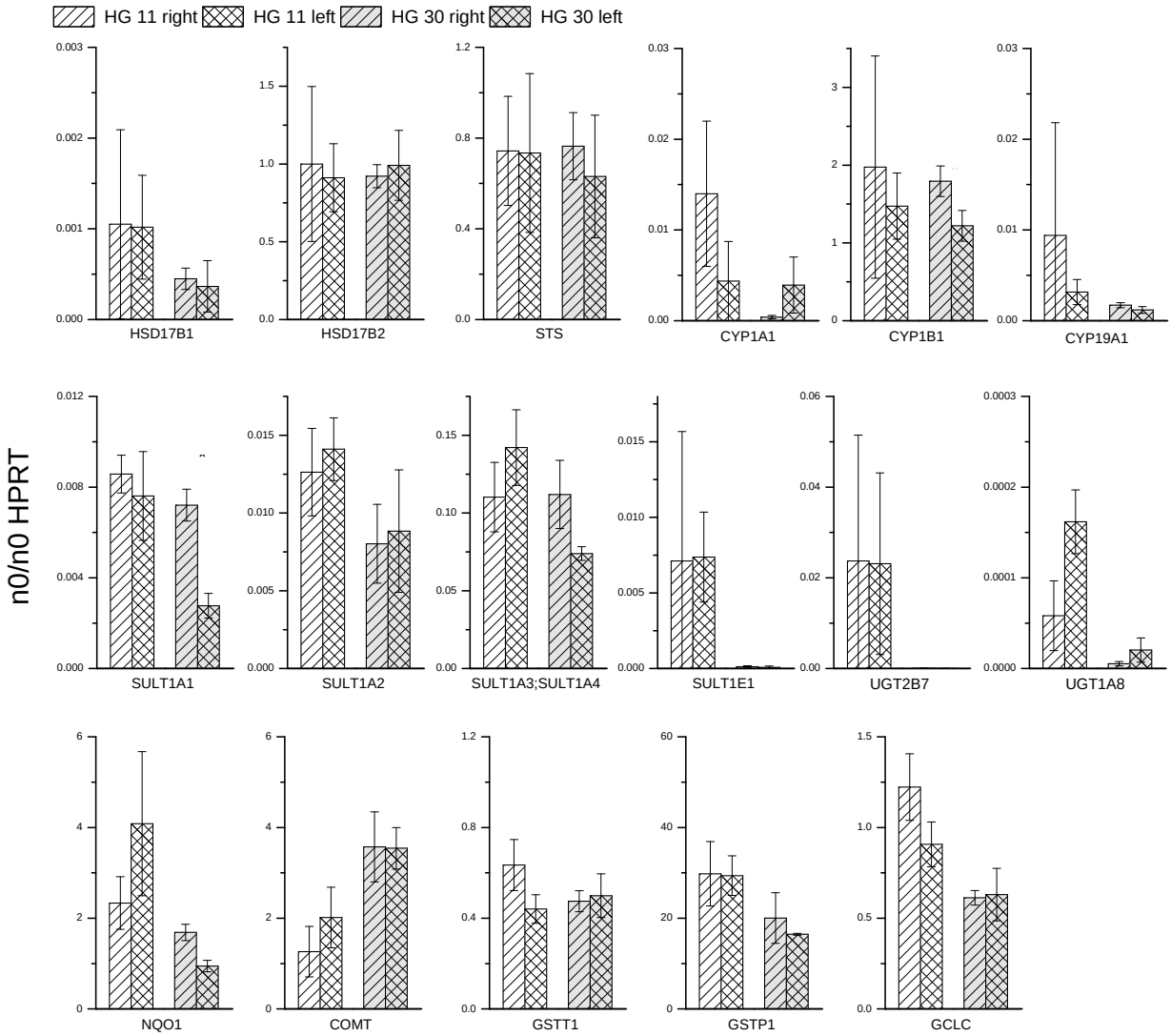


Figure 71: Comparison of transcript level of enzymes involved in E1/E2 metabolism of homogenate of three aliquots of right and left breast of sample #11 (HG 11) and sample #30 (HG 30). No significant differences (t-test, p value adjustment: Holm).

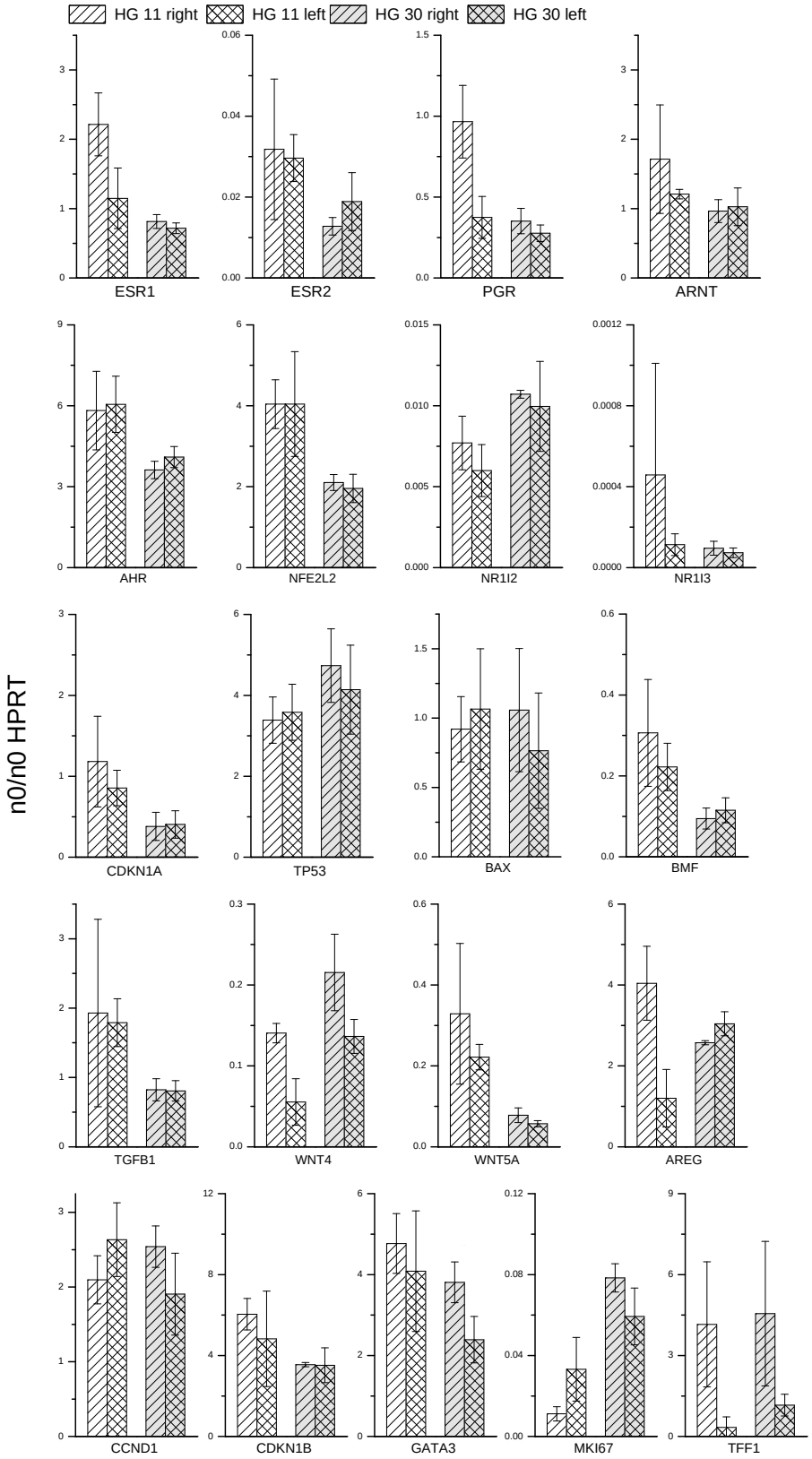


Figure 72: Comparison of transcript levels of nuclear receptors, genes involved in cell cycle, apoptosis and paracrine signal transduction, marker genes for proliferation of homogenate (HG) of three aliquots of right and left breast of sample #11 (HG 11) and sample #30 (HG 30). No significant differences (t-test, p value adjustment: Holm).

9.3.3 Calculation of expected ΔCt values

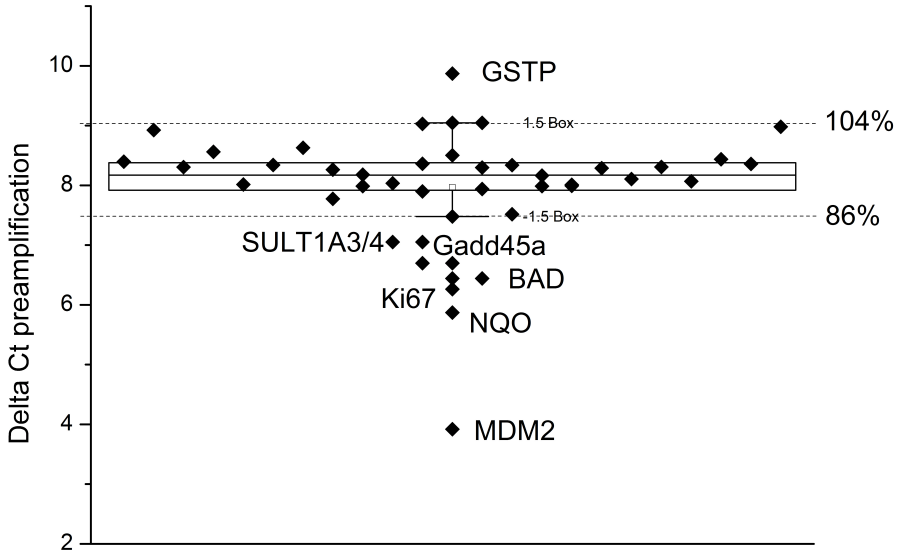


Figure 73: Box plot of preamplification experiment II (12 cycles of preamplification). The nominal ΔCt value was 8.7, whiskers of the box plot were at 7.5 and 9.0 which equals 86% and 104%, respectively.

9.4 Transcript levels of homogenate

Table 61: Concentration of isolated RNA (ng/μl, total volume 65 μl) from human mammary gland tissues (approx. 200 mg $\widehat{=}$ 8x10^7 cells). Mean yield of all samples was 16.8 μg RNA/200 mg tissue. *, RNA were not reversely transcribed into cDNA due to insufficient RNA yield.

HG #	RNA (ng/μl)
0	44.6
1	91.0
2	237.5
3*	27.2
	6.0
	2.6
4	140.4
5	31.8
6*	8.0
	18.2
7	64.5
8	67.8
10*	30.4
	42.2
11	147.8
	585.2
	384.6
	322.8
	227.9
	221.0
12	457.8
13	162.8
14	313.6
15*	44.8
	39.7
16	58.1
17	253.8
18	1043.8
19	66.6
20	327.3
21	193.8
24	126.5
25	184.2
26	245.1
27	260.2
28	480.7
29	1567.0
30	861.3
	636.2
	575.4
	640.4
	329.2
	679.4
31*	40.8
32	110.1

Table 62: Levels of all transcripts investigated in this work relative to HPRT (n0/n0 HPRT). Evaluated with SDS. Samples 0-18. M, mean.

Transcript	n0/n0 HPRT															
	0	1	2	4	5	7	8	M 11	SD 11	12	13	14	16	17	18
CYP1A1	8.4E-01	2.2E-03	1.5E-01	1.1E-03	5.5E-04	1.9E-03	1.1E-03	9.2E-03	8.2E-03	1.2E-04	3.0E-03	2.1E-03	1.1E-01	2.4E-03	1.6E-03
CYP1B1	7.2E-01	5.0E-01	1.2E+00	1.4E+00	8.2E-01	1.2E+00	9.8E-01	1.7E+00	1.1E+00	9.6E-01	7.5E-01	1.1E+00	9.3E-01	2.0E+00	1.2E+00
CYP19A1	8.0E-03	2.5E-03	5.7E-03	2.2E-03	2.9E-03	7.9E-03	4.3E-03	6.3E-03	9.5E-03	4.1E-04	3.1E-03	7.3E-04	3.9E-03	1.4E-02	3.2E-03
UGT2B7	7.6E-05	1.3E-02	8.0E-03	1.4E-05	2.2E-02	3.4E-04	5.4E-04	2.3E-02	2.3E-02	–	1.1E-05	1.9E-04	4.9E-04	1.5E-03	–
UGT1A8	3.7E-05	3.8E-04	7.7E-06	4.9E-05	3.2E-05	3.4E-05	1.8E-04	1.1E-04	6.7E-05	1.4E-05	–	–	4.8E-05	–	2.6E-06
UGT1A1	1.7E-04	–	–	3.7E-04	–	1.0E-04	–	–	–	–	8.5E-05	4.1E-05	2.3E-04	–	4.4E-04
UGT1A6	–	–	8.1E-05	2.3E-05	–	1.2E-04	–	4.2E-04	–	–	–	1.6E-04	–	6.8E-05	–
UGT1A7	1.3E-03	2.3E-03	1.7E-04	4.6E-04	3.2E-04	9.3E-04	1.5E-04	–	–	6.7E-04	1.6E-04	6.5E-04	5.6E-05	1.3E-04	–
UGT1A10	9.6E-05	–	2.3E-04	9.6E-04	5.2E-04	2.3E-04	9.1E-05	4.1E-04	–	4.8E-04	9.1E-05	3.3E-04	1.2E-04	–	1.7E-04
SULT1A1	8.7E-03	7.3E-03	7.0E-03	7.2E-03	2.5E-02	6.4E-03	7.6E-03	8.1E-03	1.5E-03	5.6E-03	5.3E-03	6.8E-03	1.4E-02	8.9E-03	5.5E-03
SULT1A2	2.4E-01	2.9E-02	1.5E-02	2.6E-02	4.5E-02	1.8E-02	3.4E-02	1.3E-02	2.5E-03	8.4E-03	1.3E-02	4.4E-02	4.3E-02	1.4E-02	1.8E-02
SULT1A3/4	1.7E-01	6.8E-02	1.1E-01	1.7E-01	1.4E-01	8.2E-02	1.5E-01	1.3E-01	2.8E-02	6.5E-02	8.5E-02	9.1E-02	1.6E-01	1.0E-01	1.1E-01
SULT1E1	8.8E-05	1.6E-02	5.2E-03	1.4E-05	2.9E-04	3.7E-03	5.6E-03	7.2E-03	6.3E-03	5.6E-03	4.0E-04	1.5E-03	9.2E-05	1.3E-03	5.6E-04
SULT2A1	–	–	2.2E-05	–	–	1.9E-05	–	6.1E-05	–	–	–	–	–	–	–
COMT	4.2E+00	1.6E+00	3.2E+00	6.4E-01	1.8E+00	3.3E+00	1.6E+00	1.6E+00	7.1E-01	3.5E+00	3.3E+00	2.1E+00	5.3E+00	1.7E+00	1.2E+00
GSTP1	3.9E+01	1.6E+01	3.1E+01	2.3E+01	2.8E+01	3.9E+01	1.2E+01	3.0E+01	5.2E+00	1.4E+01	1.8E+01	2.6E+01	3.8E+01	2.2E+01	2.2E+01
GSTT1	2.6E+00	5.8E-01	6.0E-01	3.3E-01	7.0E-01	7.9E-01	1.5E+00	5.4E-01	1.4E-01	3.4E-01	5.5E-01	4.8E-01	2.4E+00	8.5E-06	3.6E-01
NQO1	2.3E+01	2.2E+00	3.9E+00	1.8E+00	2.7E+00	1.5E+00	1.3E+00	3.5E+00	1.7E+00	1.1E+00	1.6E+00	2.0E+00	6.1E+00	1.9E+00	5.9E-01
HSD17B1	5.1E-02	3.6E-03	1.3E-03	6.9E-04	1.7E-02	3.6E-03	1.1E-02	1.0E-03	8.2E-04	1.2E-03	6.5E-04	3.7E-04	2.5E-02	4.3E-04	7.7E-04
HSD17B2	8.1E-02	3.7E-01	4.1E-01	2.9E-01	3.8E-01	2.6E-01	1.3E-01	9.6E-01	3.8E-01	2.5E-01	4.7E-01	3.2E-01	2.4E-01	2.1E-01	3.9E-01
STS	2.7E+00	5.7E-01	5.8E-01	4.6E-01	7.1E-01	1.3E+00	7.1E-01	7.4E-01	2.8E-01	5.1E-01	8.4E-01	5.2E-01	1.2E+00	1.0E+00	4.7E-01
ESR1	1.5E+00	1.1E+00	1.8E+00	6.6E-01	1.6E+00	3.0E+00	1.1E+00	1.7E+00	7.2E-01	6.3E-01	8.4E-01	3.0E+00	9.5E-01	1.1E+00	6.7E-01
ESR2	3.6E-01	2.9E-02	1.7E-02	1.4E-02	7.2E-02	2.8E-02	1.0E-01	3.1E-02	1.3E-02	1.2E-02	1.2E-02	1.5E-02	5.9E-02	1.7E-02	2.4E-02
PGR	3.2E-01	4.5E-01	2.6E-01	3.8E-01	1.4E-01	3.3E-01	3.1E-01	6.7E-01	3.7E-01	2.6E-01	8.3E-01	3.1E-01	4.0E-01	2.9E-01	3.0E-01
AHR	6.3E+00	3.7E+00	5.7E+00	3.6E+00	2.5E+00	6.2E+00	5.2E+00	5.9E+00	1.2E+00	2.5E+00	3.1E+00	5.6E+00	5.5E+00	5.7E+00	5.4E+00
NFE2L2	3.6E+00	2.9E+00	2.9E+00	1.5E+00	3.5E+00	3.7E+00	3.0E+00	4.0E+00	9.2E-01	2.3E+00	2.4E+00	3.2E+00	4.7E+00	3.5E+00	2.8E+00
NR1I3	–	1.6E-02	2.7E-04	2.6E-04	1.1E-04	1.9E-05	2.0E-03	2.9E-04	4.3E-04	1.6E-04	8.4E-05	1.0E-04	1.7E-04	1.9E-04	4.1E-04
NR1I2	2.8E-03	1.6E-02	7.7E-03	2.1E-02	9.4E-03	3.4E-03	1.4E-02	6.9E-03	1.8E-03	7.9E-03	1.5E-02	3.8E-03	7.9E-03	1.5E-02	1.5E-02
MKI67	5.8E-03	2.4E-02	1.6E-02	4.6E-02	1.2E-02	2.4E-02	1.6E-02	2.2E-02	1.6E-02	6.4E-02	4.2E-02	2.7E-03	1.3E-02	4.4E-02	2.7E-02
CDKN1B	5.8E+00	4.4E+00	3.2E+00	2.1E+00	3.1E+00	5.3E+00	5.9E+00	5.4E+00	1.7E+00	2.6E+00	2.8E+00	5.7E+00	7.3E+00	5.3E+00	5.9E+00
TFF1	3.1E-01	4.6E-01	1.7E+00	3.9E-01	2.9E-02	6.0E-02	1.6E-02	2.2E+00	2.7E+00	2.0E+00	5.6E+00	9.9E-01	3.5E+00	8.6E-01	3.4E-01
CCND1	2.4E+00	5.6E+00	–	1.3E+00	2.5E+00	2.8E+00	–	2.4E+00	4.9E-01	–	3.9E+00	2.1E+00	7.2E+00	1.9E+00	7.5E+00
GATA3	2.0E+00	4.3E+00	2.8E+00	2.3E+00	5.2E+00	5.9E+00	5.6E+00	4.4E+00	1.1E+00	2.1E+00	3.9E+00	6.7E+00	2.7E+00	3.6E+00	3.4E+00
AREG	3.6E-01	9.5E-01	5.8E+00	1.8E+00	5.2E-01	3.2E-01	3.6E-01	2.6E+00	1.7E+00	2.7E+00	2.8E+00	1.4E+00	1.3E+00	3.8E+00	1.0E+00
TGFB1	1.6E+00	1.2E+00	9.5E-01	8.7E-01	1.4E+00	2.6E+00	1.5E+00	1.9E+00	9.8E-01	6.7E-01	8.5E-01	2.5E+00	1.9E+00	1.8E+00	1.9E+00
TBP	3.1E-01	3.0E-01	2.7E-01	2.4E-01	2.9E-01	3.9E-01	3.9E-01	3.1E-01	7.4E-02	1.6E-01	1.7E-01	3.0E-01	3.3E-01	2.1E-01	1.9E-01
WNT5A	3.9E-02	2.3E-01	2.4E-01	7.3E-02	1.8E-01	4.2E-01	4.9E-01	2.8E-01	1.4E-01	1.1E-01	1.2E-01	2.3E-01	1.5E-01	1.2E-01	8.2E-02
WNT4	2.5E-02	1.3E-01	4.9E-01	4.0E-01	4.4E-02	1.5E-01	9.0E-02	9.8E-02	5.1E-02	3.0E-01	1.7E-01	5.0E-02	1.4E-01	2.9E-01	5.9E-02
GCLC	3.5E+00	1.1E+00	8.5E-01	6.5E-01	9.4E-01	1.2E+00	1.5E+00	1.1E+00	2.3E-01	5.7E-01	9.3E-01	1.1E+00	1.9E+00	9.0E-01	7.7E-01
TP53	1.9E+00	3.6E+00	3.5E+00	3.4E+00	2.9E+00	2.4E+00	3.2E+00	3.5E+00	6.0E-01	3.5E+00	2.8E+00	2.1E+00	3.1E+00	4.3E+00	3.0E+00
GADD45A	1.7E-01	1.7E-01	8.3E-02	6.9E-02	1.7E-01	1.2E-01	2.9E-01	1.6E-01	1.2E-01	4.4E-02	1.0E-01	1.9E-01	1.6E-01	1.0E-01	1.5E-01
CDKN1A	7.9E-01	4.6E-01	4.2E-01	1.9E-01	2.2E+00	1.4E+00	8.0E-01	1.0E+00	4.6E-01	3.2E-01	3.8E-01	1.6E+00	1.0E+00	3.9E-01	5.5E-01
SFN	–	4.2E-03	2.6E-04	2.6E-05	–	5.1E-04	3.6E-03	5.4E-04	8.7E-04	7.0E-05	4.4E-05	2.2E-04	–	8.4E-04	5.9E-04
MDM2	6.5E-02	2.1E-02	4.5E-02	1.2E-01	4.8E-02	1.4E-01	2.5E-02	6.4E-02	2.7E-02	3.5E-02	3.6E-02	4.7E-02	6.4E-02	3.8E-02	3.3E-02
BAX	6.0E-01	5.8E-01	7.3E-01	5.2E-01	8.1E-01	2.0E+00	7.0E-01	9.9E-01	3.3E-01	5.0E-01	6.9E-01	1.0E+00	8.8E-01	1.1E+00	6.5E-01
BAD	1.6E-01	1.5E-01	4.4E-02	4.3E-02	9.9E-02	1.1E-01	2.4E-01	7.2E-02	3.3E-02	4.1E-02	6.1E-02	5.0E-02	1.3E-01	5.8E-02	4.2E-02
BMF	1.9E-01	1.7E-01	9.5E-02	1.1E-01	1.8E-01	1.5E-01	3.0E-01	2.6E-01	1.1E-01	5.6E-02	6.6E-02	2.0E-01	2.3E-01	1.7E-01	2.7E-01
ARNT	2.5E+00	1.3E+00	7.5E-01	1.5E+00	1.4E+00	1.6E+00	1.7E+00	1.5E+00	6.2E-01	5.3E-01	7.7E-01	2.5E+00	1.9E+00	2.1E+00	1.6E+00
GUSB	2.4E+00	1.1E+00	3.4E+00	1.3E+00	1.4E+00	2.4E+00	1.4E+00	2.7E+00	1.1E+00	1.2E+00	1.5E+00	2.0E+00	2.5E+00	1.7E+00	1.3E+00

Table 63: Levels of all transcripts investigated in this work relative to HPRT (n0/n0 HPRT). Evaluated with SDS. Samples 19-32. M, mean.

Transcript	n0/n0 HPRT 19	20	21	24	25	26	27	28	29	M 30	SD 30	31	M	SD
CYP1A1	6.4E-04	6.2E-02	4.1E-01	2.6E-02	6.8E-04	3.8E-04	2.0E-03	4.1E-05	1.8E-03	2.2E-03	2.8E-03	1.8E-03	6.5E-02	8.5E-02
CYP1B1	5.7E-01	1.4E+00	7.2E-01	1.1E+00	1.0E+00	1.0E+00	5.8E-01	6.4E-01	1.1E+00	1.5E+00	3.6E-01	5.6E-01	1.0E+00	3.9E-01
CYP19A1	4.9E-04	1.0E-03	1.3E-03	1.6E-03	9.7E-04	5.5E-03	9.1E-04	3.6E-03	3.1E-04	1.4E-03	4.1E-04	7.9E-04	3.3E-03	3.2E-03
UGT2B7	6.2E-04	–	2.0E-04	1.0E-02	4.4E-03	–	1.4E-02	2.1E-04	5.1E-05	4.0E-05	2.4E-05	2.5E-02	5.9E-03	9.0E-03
UGT1A8	1.3E-04	1.3E-04	5.7E-06	2.4E-04	4.6E-05	–	1.6E-04	1.9E-05	7.2E-06	1.3E-05	1.2E-05	2.2E-05	7.9E-05	9.3E-05
UGT1A1	2.8E-04	3.0E-04	2.1E-04	7.1E-04	–	3.6E-04	–	3.5E-05		1.0E-04	–	4.1E-05	2.3E-04	1.9E-04
UGT1A6	1.6E-05	–	–	–	–	1.0E-04	2.1E-05	–	–	–	–	9.7E-05	1.1E-04	1.1E-04
UGT1A7	1.3E-04	9.1E-04	1.4E-03	1.7E-03	5.0E-04	–	5.6E-04	6.6E-05	8.5E-05	3.4E-04	–	–	6.2E-04	6.0E-04
UGT1A10	1.4E-04	3.6E-04	5.7E-04	1.2E-03	1.1E-04	5.3E-04	5.9E-04	6.3E-05	7.4E-05	–	–	4.7E-05	3.4E-04	3.1E-04
SULT1A1	9.5E-03	4.2E-03	3.2E-03	6.2E-03	3.9E-03	4.1E-03	4.8E-03	2.6E-03	4.0E-03	5.0E-03	2.5E-03	5.0E-03	7.0E-03	4.5E-03
SULT1A2	1.6E-02	7.8E-03	2.3E-02	1.3E-02	2.2E-02	1.2E-02	1.7E-02	5.1E-03	7.0E-03	8.4E-03	3.0E-03	1.1E-02	2.8E-02	1.2E-02
SULT1A3/4	1.9E-01	1.8E-01	5.5E-02	9.1E-02	8.5E-02	9.1E-02	1.4E-01	1.7E-01	7.9E-02	9.3E-02	2.5E-02	1.5E-01	1.2E-01	4.4E-02
SULT1E1	6.8E-03	6.2E-05	1.4E-04	9.7E-02	1.6E-03	1.1E-03	6.3E-03	4.3E-03	8.7E-05	9.7E-05	7.2E-05	1.0E-03	6.6E-03	1.9E-02
SULT2A1	–	–	–	–	–	–	–	–	–	–	–	–	3.4E-05	1.9E-05
COMT	1.4E+00	4.7E+00	3.1E+00	1.8E+00	1.8E+00	1.3E+00	1.8E+00	9.3E-01	1.8E+00	3.6E+00	5.7E-01	3.8E+00	2.4E+00	1.2E+00
GSTP1	3.1E+01	3.2E+01	3.3E+01	3.1E+01	2.4E+01	3.2E+01	2.8E+01	2.3E+01	2.3E+01	1.8E+01	4.0E+00	2.4E+01	2.6E+01	8.8E+00
GSTT1	4.2E-01	7.0E-04	5.7E-01	1.1E+00	4.8E-01	4.0E-01	4.0E-01	6.7E-01	4.7E-01	4.9E-01	6.9E-02	6.1E-01	6.9E-01	4.8E-01
NQO1	3.7E+00	1.4E+00	1.9E+00	4.9E+00	2.1E+00	1.2E+00	2.8E+00	1.2E+00	7.1E-01	1.3E+00	4.3E-01	1.7E+00	3.0E+00	1.3E+00
HSD17B1	3.5E-03	2.0E-03	4.7E-03	3.9E-03	1.6E-04	8.8E-04	5.7E-04	2.6E-04	9.5E-05	4.1E-04	2.0E-04	5.4E-03	5.6E-03	5.9E-03
HSD17B2	9.1E-02	2.6E-01	6.6E-01	2.9E-01	3.2E-01	2.3E-01	4.4E-01	9.0E-02	4.1E-01	9.6E-01	1.5E-01	1.2E+00	3.9E-01	2.7E-01
STS	4.6E-01	6.2E-01	6.3E-01	7.3E-01	5.7E-01	4.3E-01	4.8E-01	3.2E-01	3.2E-01	7.0E-01	2.1E-01	1.1E+00	7.4E-01	2.7E-01
ESR1	2.7E+00	1.5E+00	1.4E+00	9.9E-01	1.5E+00	1.5E+00	2.6E+00	6.9E-01	4.5E-01	7.7E-01	9.6E-02	1.7E+00	1.4E+00	7.6E-01
ESR2	2.3E-02	1.9E-02	2.1E-02	6.1E-02	1.0E-02	1.2E-02	1.5E-02	1.2E-02	6.6E-03	1.6E-02	5.8E-03	6.8E-02	4.2E-02	2.4E-02
PGR	1.6E-01	2.3E-01	2.3E-01	9.3E-02	9.8E-01	4.4E-01	1.5E-01	2.8E-01	2.4E-01	3.1E-01	7.2E-02	5.6E-01	3.6E-01	2.1E-01
AHR	5.5E+00	8.9E+00	4.8E+00	5.0E+00	5.4E+00	6.3E+00	5.7E+00	5.9E+00	2.3E+00	3.9E+00	4.1E-01	6.0E+00	5.1E+00	1.8E+00
NFE2L2	3.2E+00	3.4E+00	2.9E+00	3.5E+00	2.7E+00	2.0E+00	2.4E+00	2.5E+00	1.0E+00	2.0E+00	2.7E-01	3.3E+00	2.9E+00	1.0E+00
NR1I3	5.1E-05	6.4E-05	1.5E-04	2.5E-04	7.6E-05	2.0E-04	5.6E-05	8.2E-04	8.1E-05	8.4E-05	2.9E-05	9.4E-05	9.1E-04	3.1E-03
NR1I2	6.1E-03	1.2E-02	1.3E-02	9.3E-03	8.4E-03	7.1E-03	9.7E-03	6.6E-03	1.5E-02	1.0E-02	1.8E-03	1.1E-02	1.0E-02	4.8E-03
MKI67	7.9E-03	1.5E-02	1.2E-02	2.1E-02	4.3E-02	1.3E-02	1.9E-02	2.2E-02	8.6E-02	6.9E-02	1.4E-02	1.4E-02	2.7E-02	1.9E-02
CDKN1B	3.6E+00	7.1E+00	4.8E+00	4.9E+00	3.4E+00	3.2E+00	4.7E+00	3.2E+00	1.5E+00	3.5E+00	5.5E-01	4.5E+00	4.4E+00	1.7E+00
TFF1	2.2E-01	2.8E-02	2.7E-01	3.5E-01	2.0E+01	5.2E-01	4.8E-02	1.1E+00	1.1E+00	2.9E+00	2.5E+00	7.5E-01	1.8E+00	3.9E+00
CCND1	–	5.7E+00	5.2E+00	7.9E-01	1.5E+00	1.2E+00	–	8.2E-01	1.2E+00	2.2E+00	5.2E-01	2.8E+00	3.1E+00	2.1E+00
GATA3	7.7E+00	4.8E+00	3.2E+00	3.3E+00	2.8E+00	3.6E+00	5.1E+00	4.4E+00	1.6E+00	3.1E+00	9.1E-01	5.6E+00	4.0E+00	1.6E+00
AREG	2.6E-01	1.7E+00	1.8E+00	2.3E-01	4.7E+00	7.0E-01	6.5E-01	2.4E+00	1.3E+00	2.8E+00	3.2E-01	1.3E+00	1.7E+00	1.4E+00
TGFB1	1.3E+00	1.7E+00	1.0E+00	1.6E+00	9.7E-01	1.5E+00	1.5E+00	1.1E+00	4.6E-01	8.1E-01	1.4E-01	1.2E+00	1.4E+00	5.6E-01
TBP	1.8E-01	2.9E-01	2.7E-01	2.1E-01	2.5E-01	2.0E-01	2.2E-01	1.8E-01	1.4E-01	2.5E-01	5.5E-02	4.4E-01	2.6E-01	9.0E-02
WNT5A	2.1E-01	6.9E-02	9.9E-02	2.9E-01	1.2E-01	7.1E-02	1.1E-01	2.5E-01	4.0E-02	6.8E-02	1.7E-02	1.9E-01	1.7E-01	1.1E-01
WNT4	6.0E-02	8.4E-02	1.2E-01	5.8E-02	9.9E-02	5.0E-02	4.9E-02	1.6E-01	3.1E-01	1.8E-01	5.4E-02	2.8E-01	1.6E-01	1.2E-01
GCLC	9.3E-01	1.0E+00	1.0E+00	1.2E+00	6.7E-01	8.0E-01	1.0E+00	4.8E-01	4.8E-01	6.2E-01	9.5E-02	1.3E+00	1.1E+00	3.7E-01
TP53	1.9E+00	3.4E+00	3.1E+00	2.5E+00	2.9E+00	2.9E+00	2.4E+00	2.4E+00	4.0E+00	4.4E+00	9.6E-01	3.7E+00	3.1E+00	8.9E-01
GADD45A	1.1E-01	7.2E-02	9.4E-02	1.6E-01	7.3E-02	6.9E-02	3.8E-02	3.9E-02	4.2E-02	5.4E-02	2.3E-02	9.0E-02	1.1E-01	6.0E-02
CDKN1A	1.0E+00	4.5E-01	2.8E-01	8.9E-01	3.6E-01	4.6E-01	5.9E-01	6.5E-01	2.0E-01	3.9E-01	1.5E-01	7.4E-01	7.0E-01	4.8E-01
SFN	3.3E-04	2.3E-04	5.4E-05	4.7E-05	–	2.2E-04	–	–	–	1.7E-04	1.2E-04	1.1E-04	6.7E-04	1.1E-03
MDM2	3.9E-02	5.9E-02	3.6E-02	3.2E-02	4.3E-02	2.6E-02	3.8E-02	3.8E-02	2.9E-02	6.6E-02	2.5E-02	1.2E-01	5.2E-02	3.1E-02
BAX	5.9E-01	7.2E-01	7.5E-01	7.1E-01	6.1E-01	5.7E-01	6.0E-01	6.6E-01	4.2E-01	9.1E-01	4.2E-01	1.1E+00	7.7E-01	3.2E-01
BAD	5.3E-02	6.2E-02	4.3E-02	5.2E-02	7.4E-02	5.4E-02	5.1E-02	3.5E-02	3.4E-02	4.1E-02	1.2E-02	7.8E-02	7.5E-02	4.6E-02
BMF	1.4E-01	3.0E-01	1.5E-01	1.2E-01	1.4E-01	1.4E-01	1.5E-01	1.0E-01	3.9E-02	1.0E-01	2.8E-02	1.5E-01	1.6E-01	7.2E-02
ARNT	2.0E+00	1.3E+00	1.1E+00	2.1E+00	1.7E+00	1.8E+00	1.2E+00	1.3E+00	5.8E-01	1.0E+00	2.0E-01	1.3E+00	1.5E+00	5.3E-01
GUSB	1.3E+00	1.7E+00	1.9E+00	3.0E+00	2.3E+00	1.2E+00	2.4E+00	1.2E+00	1.1E+00	1.4E+00	1.9E-01	2.9E+00	1.9E+00	7.3E-01

Table 64: Levels of all transcripts investigated in this work relative to HPRT (n0/n0 HPRT). Evaluated with ExpressionSuite software. Samples 0-20.

	N0/n0 HPRT Sample #															
Transcripts	0	1	2	4	5	7	8	11	12	13	14	16	17	18	19	20
Apoptosis																
BAD	1.5E-01	1.7E-01	3.9E-02	3.1E-02	1.5E-01	1.1E-01	3.4E-01	7.1E-02	3.1E-02	1.1E-01	5.0E-02	1.4E-01	5.3E-02	3.7E-02	4.7E-02	4.9E-02
BMF	1.9E-01	1.4E-01	7.7E-02	2.3E-01	2.2E-01	1.4E-01	7.0E-01	2.4E-01	3.5E-02	4.7E-02	1.3E-01	1.9E-01	1.5E-01	2.5E-01	7.9E-02	8.5E-01
BAX	2.1E+00	8.3E-01	1.1E+00	1.3E+00	1.1E+00	2.9E+00	1.3E+00	3.7E+00	6.6E-01	1.0E+00	1.7E+00	1.8E+00	1.8E+00	7.2E-01	6.8E-01	5.1E+00
Cell cycle																
CDKN1A	1.1E+00	6.6E-01	4.5E-01	1.7E-01	3.9E+00	1.4E+00	1.4E+00	1.1E+00	3.3E-01	3.6E-01	1.2E+00	1.1E+00	2.4E-01	3.6E-01	6.0E-01	5.6E-01
CDKN1B	6.6E+00	3.5E+00	3.0E+00	2.2E+00	4.7E+00	5.0E+00	5.7E+00	4.3E+00	1.9E+00	2.1E+00	3.7E+00	7.3E+00	3.6E+00	3.5E+00	2.0E+00	6.4E+00
GATA3	2.5E+00	5.2E+00	3.1E+00	2.1E+00	6.1E+00	5.7E+00	7.2E+00	4.3E+00	2.3E+00	3.9E+00	5.4E+00	3.3E+00	2.6E+00	2.7E+00	5.4E+00	5.3E+00
GADD45A	1.7E-01	1.5E-01	5.8E-02	5.3E-02	1.8E-01	1.2E-01	3.1E-01	1.2E-01	4.2E-02	7.1E-02	1.1E-01	1.4E-01	5.5E-02	8.5E-02	5.1E-02	5.0E-02
MDM2	8.2E-02	3.1E-02	8.0E-02	1.3E-01	5.8E-02	1.2E-01	9.4E-02	8.2E-02	1.2E-01	3.5E-02	5.0E-02	8.5E-02	7.5E-02	3.1E-02	4.7E-02	7.3E-02
TP53	1.8E+00	3.0E+00	2.8E+00	2.6E+00	2.2E+00	1.7E+00	2.7E+00	2.6E+00	2.7E+00	2.1E+00	1.6E+00	2.8E+00	2.9E+00	2.2E+00	1.5E+00	2.6E+00
SFN	0.0E+00	1.90E-03	4.73E-04	2.64E-05	0.0E+00	1.3E-02	2.2E-02	8.1E-04	7.2E-05	5.4E-05	8.0E-05	0.0E+00	2.9E-04	1.2E-03	9.7E-05	2.0E-03
Marker proliferation and ESR1 activation																
MKI67	4.8E-03	2.0E-02	1.2E-02	3.3E-02	1.7E-02	2.6E-02	1.9E-02	1.6E-02	4.6E-02	3.4E-02	1.9E-03	1.1E-02	2.8E-02	1.8E-02	4.4E-03	1.0E-02
CCND1	4.3E+00	7.3E+00	0.0E+00	1.8E+00	3.3E+00	4.2E+00	0.0E+00	6.1E+00	0.0E+00	4.2E+00	2.2E+00	1.1E+01	1.7E+00	3.0E+00	0.0E+00	1.5E+01
TFF1	4.0E-01	3.4E-01	1.7E+00	6.9E-01	7.4E-02	8.6E-02	1.8E-02	2.1E+00	1.7E+00	5.1E+00	1.5E+00	4.0E+00	5.4E-01	1.8E-01	1.1E-01	7.2E-01
E1/E2 metabolism																
HSD17B1	8.6E-02	1.1E-02	2.7E-03	2.1E-03	4.0E-02	7.9E-03	2.3E-02	2.4E-03	1.6E-03	1.1E-03	8.6E-04	3.0E-02	9.1E-04	1.5E-03	4.3E-03	9.7E-03
HSD17B2	1.1E-01	4.2E-01	5.7E-01	3.6E-01	4.9E-01	2.3E-01	1.7E-01	1.2E+00	2.0E-01	3.9E-01	4.8E-01	2.0E-01	2.1E-01	4.2E-01	8.3E-02	7.3E-01
COMT	4.5E+00	5.8E-01	4.7E-01	2.2E-01	5.6E-01	5.9E-01	6.9E-01	3.4E-01	7.7E-01	4.8E-01	1.8E+00	1.1E+00	6.3E-01	4.3E-01	1.1E+00	6.0E+00
CYP1A1	2.0E+00	3.1E-03	1.5E-01	7.3E-02	6.4E-04	4.8E-03	3.6E-03	1.5E-02	1.0E-04	1.6E-02	1.0E-02	2.1E-01	2.6E-03	2.5E-03	4.3E-04	4.9E-02
CYP1B1	8.1E-01	4.0E-01	9.6E-01	3.1E+00	1.1E+00	9.7E-01	9.3E-01	1.5E+00	8.4E-01	6.8E-01	9.5E-01	8.7E-01	1.4E+00	1.0E+00	3.8E-01	2.4E+00
CYP19A1	2.1E-02	2.5E-03	8.1E-03	8.5E-03	1.1E-02	1.7E-02	4.6E-03	1.3E-02	7.0E-04	4.4E-03	1.5E-03	5.6E-03	2.3E-02	7.7E-03	8.2E-04	1.4E-03
GCLC	4.7E+00	7.7E-01	7.8E-01	5.9E-01	1.1E+00	1.0E+00	1.5E+00	1.1E+00	4.3E-01	8.5E-01	8.1E-01	1.7E+00	5.8E-01	5.1E-01	6.3E-01	1.1E+00
GSTP1	3.9E+01	1.6E+01	3.1E+01	2.3E+01	2.8E+01	3.9E+01	1.2E+01	3.0E+01	1.4E+01	1.8E+01	2.6E+01	3.8E+01	2.2E+01	2.2E+01	3.1E+01	3.2E+01
GSTT1	3.2E+00	5.0E-01	4.7E-01	3.3E-01	5.6E-01	7.3E-01	1.5E+00	5.4E-01	2.3E-01	4.1E-01	4.4E-01	1.8E+00	6.8E-06	3.4E-01	3.6E-01	5.9E-04
NQO1	2.3E+01	2.2E+00	3.9E+00	1.8E+00	2.7E+00	1.5E+00	1.3E+00	3.5E+00	1.1E+00	1.6E+00	2.0E+00	6.1E+00	1.9E+00	5.9E-01	3.7E+00	1.4E+00
STS	2.8E+01	7.8E-01	6.1E-01	5.2E-01	8.3E-01	1.1E+00	7.8E-01	9.5E-01	3.5E-01	8.8E-01	7.3E-01	1.1E+00	5.1E+00	2.8E+00	4.1E-01	2.3E+01
SULT1E1	3.8E-04	1.6E-02	7.1E-03	1.2E-04	9.3E-04	5.9E-03	7.1E-03	1.8E-02	8.2E-03	7.3E-04	2.4E-03	8.4E-04	1.7E-03	1.8E-03	7.5E-03	6.5E-04
SULT1A1	1.3E-02	9.5E-03	9.6E-03	7.4E-03	1.2E-01	8.9E-03	2.5E-02	1.2E-02	6.5E-03	5.9E-03	8.1E-03	3.4E-02	8.2E-03	6.1E-03	1.6E-02	5.4E-03
SULT1A2	2.2E-01	2.4E-02	1.6E-02	1.9E-02	9.1E-02	2.6E-02	4.2E-02	3.1E-02	1.0E-02	1.5E-02	3.9E-02	5.5E-02	1.1E-02	1.6E-02	9.6E-03	8.7E-03
SULT1A3/4	2.4E-01	6.1E-02	8.6E-02	1.1E-01	1.9E-01	1.0E-01	2.1E-01	9.2E-02	4.9E-02	6.9E-02	1.2E-01	1.3E-01	9.3E-02	2.4E-01	1.8E-01	5.3E-01
SULT2A1	0.0E+00	0.0E+00	6.8E-03	0.0E+00	0.0E+00	2.9E-02	0.0E+00	3.9E-04	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
UGT1A4/3	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
UGT1A8	2.3E-04	4.2E-04	1.1E-05	4.7E-04	1.8E-03	1.6E-04	3.0E-04	6.3E-04	1.8E-05	0.0E+00	0.0E+00	2.2E-04	0.0E+00	2.8E-04	2.6E-04	1.4E-01
UGT2B7	5.3E-05	3.6E-01	2.0E-02	1.4E-05	9.6E-02	4.5E-03	7.5E-04	9.0E-01	0.0E+00	1.0E-05	2.7E-04	1.8E-03	8.8E-03	0.0E+00	8.0E-03	0.0E+00
Nuclear receptors																
AHR	6.0E+00	3.2E+00	4.9E+00	4.1E+00	3.2E+00	5.5E+00	5.2E+00	4.8E+00	2.3E+00	3.1E+00	4.0E+00	4.7E+00	4.3E+00	4.1E+00	3.9E+00	7.7E+00
ARNT	2.4E+00	1.9E+00	1.0E+00	1.4E+00	1.9E+00	2.2E+00	7.6E+00	1.7E+00	7.8E-01	9.7E-01	1.9E+00	2.8E+00	1.8E+00	9.5E-01	1.8E+00	2.3E+00
ESR1	1.7E+00	1.9E+00	1.7E+00	6.8E-01	2.1E+00	2.3E+00	2.8E+00	2.4E+00	5.2E-01	6.6E-01	3.8E+00	7.6E-01	1.2E+00	7.5E-01	2.0E+00	2.9E+00
ESR2	4.2E-01	3.0E-02	1.3E-02	1.3E-02	6.2E-02	2.0E-02	1.9E-01	3.7E-02	8.2E-03	8.8E-03	1.2E-02	4.6E-02	1.2E-02	1.7E-02	1.4E-02	1.4E-02
NFE2L2	4.7E+00	2.5E+00	2.1E+00	1.6E+00	3.2E+00	4.2E+00	3.3E+00	4.4E+00	1.9E+00	2.0E+00	3.1E+00	3.7E+00	2.9E+00	3.5E+00	2.3E+00	2.4E+00
NR1I2	3.4E-03	1.3E-02	7.7E-03	2.1E-02	1.2E-02	3.2E-03	1.3E-02	6.1E-03	8.3E-03	1.4E-02	2.8E-03	8.1E-03	1.0E-02	1.2E-02	3.8E-03	1.2E-02
NR1I3	0.0E+00	8.7E-04	3.4E-04	1.6E-04	2.1E-04	2.8E-05	1.2E-04	1.4E-04	1.6E-04	7.3E-05	9.0E-05	4.2E-04	1.7E-04	4.7E-04	1.2E-04	1.3E-04
PGR	3.0E-01	4.0E-01	2.6E-01	3.2E-01	1.7E-01	2.7E-01	4.5E-01	8.9E-01	2.4E-01	8.6E-01	2.3E-01	4.4E-01	2.0E-01	2.5E-01	1.5E-01	2.1E-01
Paracrine signal transduction																
AREG	3.8E-01	6.3E-01	5.4E+00	1.5E+00	7.3E-01	2.9E-01	3.2E-01	2.5E+00	1.9E+00	2.3E+00	1.2E+00	2.4E+00	3.4E+00	7.5E-01	1.8E-01	2.2E+00
TGFB1	2.4E+00	2.9E+00	1.3E+00	1.3E+00	1.7E+00	4.5E+00	6.5E+00	3.8E+00	7.4E-01	1.1E+00	4.0E+00	3.5E+00	1.2E+00	1.5E+00	7.1E-01	4.5E+00
WNT4	2.2E-02	9.5E-02	4.0E-01	3.0E-01	6.6E-02	1.1E-01	7.4E-02	8.2E-02	2.3E-01	1.3E-01	4.8E-02	1.3E-01	2.3E-01	5.0E-02	4.7E-02	7.5E-02
WNT5A	3.6E-02	1.2E-01	1.9E-01	6.9E-02	2.0E-01	3.3E-01	3.1E-01	2.0E-01	8.8E-02	1.2E-01	1.6E-01	1.3E-01	7.5E-02	5.9E-02	1.3E-01	5.3E-02

Table 65: Levels of all transcripts investigated in this work relative to HPRT (n0/n0 HPRT). Evaluated with ExpressionSuite software. Samples 21-32 and mean of MCF-7 cells cultured with normal culture media (DMEM, n=2) or steroid-free media (SF, n=2).

	N0/n0 HPRT											
	Sample #										MCF-7	
Transcripts	21	24	25	26	27	28	29	30	32	Median	DMEM	SF
Apoptosis												
BAD	3.4E-02	4.8E-02	1.5E-01	9.2E-02	6.5E-02	3.0E-02	2.9E-02	3.8E-02	5.6E-02	5.3E-02	5.2E-02	1.2E-01
BMF	9.4E-02	7.8E-02	1.0E-01	9.0E-02	1.0E-01	9.2E-02	4.3E-02	7.4E-02	1.2E-01	1.2E-01	2.6E-03	5.4E-02
BAX	1.5E+00	1.0E+00	1.5E+00	8.5E-01	9.0E-01	6.4E+00	9.6E-01	1.1E+00	1.2E+00	1.2E+00	1.4E-01	3.3E-01
Cell cycle												
CDKN1A	2.3E-01	6.0E-01	3.9E-01	3.0E-01	3.9E-01	7.2E-01	2.3E-01	3.1E-01	5.0E-01	5.0E-01	4.3E-01	1.6E+00
CDKN1B	3.9E+00	3.3E+00	4.1E+00	2.2E+00	3.3E+00	4.3E+00	1.9E+00	3.6E+00	3.4E+00	3.6E+00	5.3E-01	2.4E+00
GATA3	2.9E+00	2.3E+00	3.3E+00	2.6E+00	3.6E+00	5.2E+00	1.9E+00	2.9E+00	5.6E+00	3.3E+00	5.0E+00	1.0E+01
GADD45A	6.0E-02	1.3E-01	6.5E-02	3.9E-02	2.4E-02	3.4E-02	3.6E-02	4.2E-02	6.0E-02	6.0E-02	2.6E-02	1.2E-01
MDM2	3.5E-02	3.1E-02	5.5E-02	3.1E-02	7.3E-02	5.2E-02	3.3E-02	8.0E-02	8.4E-02	7.3E-02	3.1E-01	2.6E-01
TP53	2.1E+00	1.9E+00	2.7E+00	2.2E+00	1.9E+00	2.2E+00	3.7E+00	3.1E+00	2.1E+00	2.2E+00	1.2E+00	2.8E+00
SFN	1.6E-04	1.6E-05	0.0E+00	8.0E-05	0.0E+00	0.0E+00	0.0E+00	1.4E-04	8.6E-05	8.0E-05	0.0E+00	0.0E+00
Marker proliferation and ESR1 activation												
MKI67	7.7E-03	1.4E-02	3.3E-02	9.3E-03	1.5E-02	1.8E-02	8.4E-02	5.5E-02	1.1E-02	1.7E-02	2.5E-02	7.4E-03
CCND1	3.8E+00	8.1E-01	3.0E+00	1.2E+00	0.0E+00	1.4E+00	2.4E+00	2.8E+00	2.2E+00	2.4E+00	6.9E-01	1.6E+00
TFF1	1.0E+00	2.1E-01	4.6E+01	2.9E-01	4.3E-02	1.5E+00	2.4E+00	1.5E+01	8.5E-01	7.2E-01	9.1E+00	1.1E+00
E1/E2 metabolism												
HSD17B1	5.5E-03	9.0E-03	8.0E-04	2.0E-03	1.2E-03	1.0E-03	4.0E-04	1.4E-03	7.8E-03	2.4E-03	0.0E+00	0.0E+00
HSD17B2	5.7E-01	3.2E-01	3.6E-01	2.3E-01	5.4E-01	9.6E-02	6.4E-01	8.1E-01	1.3E+00	3.9E-01	0.0E+00	0.0E+00
COMT	5.7E-01	7.7E-01	7.9E-01	4.8E-01	1.0E+00	2.9E-01	5.3E-01	7.8E-01	6.0E-01	6.0E-01	1.2E-01	4.3E-01
CYP1A1	2.7E-01	3.8E-01	4.9E-03	6.0E-04	9.9E-02	5.1E-04	6.5E-03	2.3E-02	5.9E-02	1.0E-02	1.4E-03	5.9E-02
CYP1B1	7.0E-01	8.8E-01	2.2E+00	7.9E-01	4.6E-01	7.3E-01	1.3E+00	1.4E+00	4.4E-01	9.3E-01	2.2E-01	6.4E-01
CYP19A1	1.4E-03	3.5E-03	4.0E-03	1.0E-02	1.9E-03	9.1E-03	7.2E-04	4.0E-03	2.2E-03	4.4E-03	0.0E+00	0.0E+00
GCLC	7.1E-01	8.9E-01	7.8E-01	5.0E-01	8.0E-01	5.2E-01	4.8E-01	5.4E-01	9.3E-01	7.8E-01	1.7E-01	7.1E-01
GSTP1	3.3E+01	3.1E+01	2.4E+01	3.2E+01	2.8E+01	2.3E+01	2.3E+01	1.8E+01	2.5E+01	2.6E+01	0.0E+00	0.0E+00
GSTT1	3.6E-01	1.1E+00	6.7E-01	3.9E-01	4.0E-01	7.9E-01	5.0E-01	4.2E-01	5.3E-01	4.7E-01	1.2E-01	7.1E-01
NQO1	1.9E+00	4.9E+00	2.1E+00	1.2E+00	2.8E+00	1.2E+00	7.1E-01	1.3E+00	3.8E+00	1.9E+00	8.3E-01	4.1E-01
STS	1.5E+00	1.8E+00	2.5E+00	3.1E-01	6.0E-01	5.4E-01	5.2E-01	1.3E+00	6.9E-01	8.3E-01	8.7E-03	2.2E-01
SULT1E1	2.4E-04	1.9E-01	7.1E-03	2.3E-03	3.8E-02	7.7E-03	1.8E-04	1.7E-04	1.2E-03	2.3E-03	0.0E+00	0.0E+00
SULT1A1	4.3E-03	1.2E-02	7.7E-03	5.7E-03	5.8E-03	4.7E-03	5.3E-03	7.9E-03	6.1E-03	7.9E-03	0.0E+00	0.0E+00
SULT1A2	2.6E-02	1.0E-02	3.2E-02	1.6E-02	2.5E-02	6.7E-03	7.2E-03	1.2E-02	1.2E-02	1.6E-02	2.3E-03	3.2E-03
SULT1A3/4	4.1E-02	8.4E-02	7.3E-02	1.0E-01	1.6E-01	3.2E-01	6.9E-02	8.2E-02	1.7E-01	1.0E-01	5.6E-02	8.2E-02
SULT2A1	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	1.1E-05	8.3E-05
UGT1A4/3	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
UGT1A8	9.9E-06	3.3E-04	6.9E-04	0.0E+00	2.4E-04	1.3E-04	3.6E-05	4.5E-04	4.1E-05	2.3E-04	0.0E+00	0.0E+00
UGT2B7	9.3E-03	1.1E-02	3.6E-03	0.0E+00	2.4E-02	4.6E-03	4.8E-05	3.9E-05	2.3E-01	3.6E-03	0.0E+00	0.0E+00
Nuclear receptors												
AHR	4.5E+00	4.1E+00	7.5E+00	5.0E+00	5.7E+00	5.4E+00	2.4E+00	3.4E+00	5.0E+00	4.5E+00	2.1E-01	1.7E+00
ARNT	1.6E+00	1.2E+00	1.5E+00	1.6E+00	9.5E-01	1.2E+00	5.3E-01	1.1E+00	1.4E+00	1.6E+00	1.8E-01	6.8E-01
ESR1	1.5E+00	9.7E-01	2.2E+00	2.6E+00	4.4E+00	8.7E-01	5.1E-01	6.1E-01	1.2E+00	1.7E+00	7.2E-01	3.3E+00
ESR2	1.4E-02	4.8E-02	9.6E-03	8.7E-03	1.4E-02	1.2E-02	5.9E-03	1.2E-02	4.3E-02	1.4E-02	0.0E+00	0.0E+00
NFE2L2	2.0E+00	4.1E+00	3.6E+00	1.7E+00	2.4E+00	3.0E+00	1.2E+00	2.3E+00	3.0E+00	2.9E+00	2.8E-01	1.4E+00
NR1I2	1.3E-02	7.7E-03	9.3E-03	5.4E-03	8.2E-03	7.2E-03	1.6E-02	8.9E-03	9.8E-03	8.9E-03	0.0E+00	0.0E+00
NR1I3	5.3E-04	2.4E-04	1.6E-04	2.2E-04	5.4E-05	9.2E-04	7.5E-05	9.3E-05	1.3E-04	1.6E-04	0.0E+00	0.0E+00
PGR	2.4E-01	8.9E-02	9.9E-01	3.5E-01	1.4E-01	3.0E-01	2.0E-01	2.5E-01	4.3E-01	2.6E-01	1.7E-01	3.2E-01
Paracrine signal transduction												
AREG	1.7E+00	1.6E-01	4.6E+00	5.6E-01	5.9E-01	2.2E+00	1.3E+00	2.2E+00	1.1E+00	1.3E+00	4.3E-02	4.3E-01
TGFB1	3.7E+00	1.1E+00	2.4E+00	1.4E+00	1.6E+00	2.0E+00	7.0E-01	1.3E+00	1.6E+00	1.6E+00	7.1E-02	3.2E-01
WNT4	9.0E-02	5.0E-02	8.9E-02	4.3E-02	4.7E-02	1.5E-01	2.9E-01	1.2E-01	2.1E-01	9.0E-02	3.2E-03	6.4E-02
WNT5A	8.5E-02	2.1E-01	1.4E-01	5.2E-02	8.4E-02	2.4E-01	3.2E-02	5.9E-02	1.3E-01	1.2E-01	8.7E-04	1.7E-02

9.5 Transcript level of epithelial cells

Table 66: n0/n0 HPRT of pre-preamplified (14+10 cycles) cDNA MECs+GECs.

	n0/n0 HPRT												
Transcript	**1**	**12**	**25**	**26**	**28**	**Mean29**	**30**	**Mean**	**SD**	**29I**	**29II**	**SD**	**rel SD**
AHR	3.85	2.02	2.83	2.20	1.74	4.33	2.53	2.8	1.0	5.69	2.98	1.9	44.3
AREG	6.12	0.38	1.73	0.34	1.57	2.66	2.09	2.1	2.0	1.80	3.52	1.2	45.5
ARNT	1.56	0.42	1.19	0.44	0.44	1.47	0.67	0.9	0.5	1.62	1.33	0.2	13.5
BAX	2.93	1.79	1.95	0.74	0.86	4.44	0.85	1.9	1.4	6.07	2.80	2.3	52.1
BMF	0.00	0.06	0.08	0.08	0.00	0.11	0.04	0.1	0.0	0.21	0.00	0.2	141.4
CCND1	5.33	0.00	2.59	0.49	2.27	12.10	5.22	4.0	4.1	18.75	5.44	9.4	77.8
CDKN1A	0.19	0.08	1.10	0.99	0.07	0.34	0.00	0.4	0.5	0.54	0.15	0.3	79.7
CDKN1B	13.89	2.20	3.29	1.16	3.60	10.36	5.97	5.8	4.7	8.63	12.09	2.4	23.6
COMT	0.68	0.35	0.36	0.79	0.34	2.74	0.86	0.9	0.9	2.13	3.36	0.9	31.6
CYP1B1	0.34	0.17	0.71	0.00	1.14	2.77	1.15	0.9	0.9	2.65	2.89	0.2	6.2
ESR1	5.85	2.37	1.44	1.49	1.92	3.81	1.15	2.6	1.7	5.08	2.55	1.8	47.0
GATA3	36.00	3.77	10.17	4.99	4.10	19.74	10.14	12.7	11.7	23.83	15.64	5.8	29.4
GCLC	1.17	0.27	1.31	0.71	0.94	2.13	0.90	1.1	0.6	1.02	3.24	1.6	73.7
GSTT1	0.24	0.27	0.32	0.07	0.51	0.83	0.32	0.4	0.2	0.95	0.71	0.2	21.1
GUSB	1.32	0.56	1.43	0.26	0.99	0.91	0.96	0.9	0.4	1.33	0.48	0.6	67.0
HSD17B2	0.00	0.48	1.01	0.00	0.54	1.73	1.04	0.7	0.6	2.42	1.05	1.0	55.7
MKI67	0.00	0.04	0.00	0.00	0.00	0.00	0.09	0.0	0.0	0.00	0.00	0.0	-
NFE2L2	3.40	1.34	3.63	0.55	1.35	6.40	1.78	2.6	2.0	9.27	3.52	4.1	63.6
NR1I2	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.00	0.00	0.0	-
PGR	1.56	0.13	1.12	0.15	0.31	0.21	0.40	0.6	0.6	0.15	0.28	0.1	41.5
SFN	0.00	0.60	0.00	0.00	0.00	0.00	0.00	0.1	0.2	0.00	0.00	0.0	-
SULT1A3/4	0.12	0.04	0.05	0.11	0.15	0.23	0.08	0.1	0.1	0.26	0.19	0.0	20.8
STS	0.87	0.09	0.40	0.29	0.43	0.20	0.10	0.3	0.3	0.24	0.17	0.0	21.6
TBP	0.33	0.15	0.17	0.07	0.13	0.17	0.20	0.2	0.1	0.13	0.22	0.1	39.8
TFF1	1.31	0.15	0.11	0.48	1.74	1.46	1.07	0.9	0.7	0.87	2.04	0.8	56.8
TGFB1	1.02	0.06	0.47	0.07	0.13	1.41	0.24	0.5	0.5	1.84	0.98	0.6	43.4
TP53	5.29	2.34	4.56	2.08	4.80	20.28	3.88	6.2	6.3	19.85	20.72	0.6	3.0
WNT4	0.61	0.31	0.21	0.00	0.09	1.35	0.31	0.4	0.5	0.96	1.73	0.5	40.4
WNT5A	0.10	0.07	0.22	0.00	0.12	0.00	0.00	0.1	0.1	0.00	0.00	0.0	-

Cluster analysis

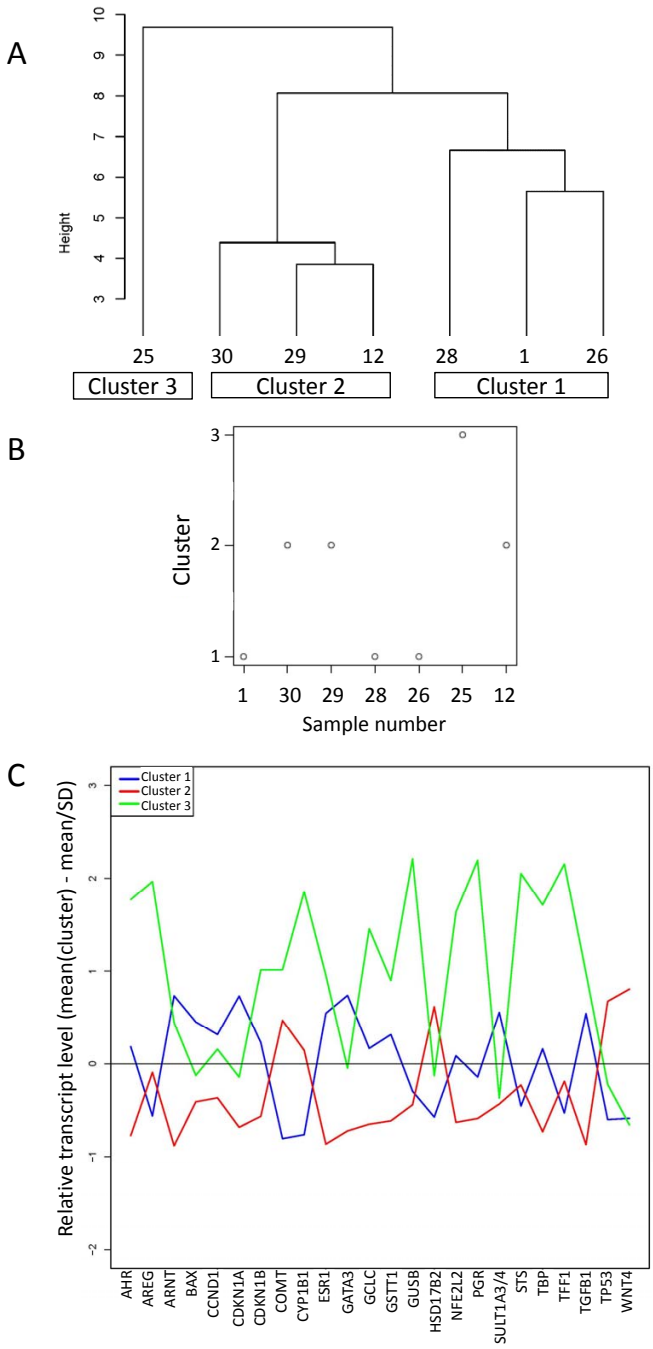


Figure 74: Dendrogram (A), cluster classification (B) and characteristics of the two main clusters concerning their relative (n0/n0 HPRT) transcript levels presented as mean(cluster)-mean(all)/SD(all) (C) of the cluster analysis of transcripts levels of all seven **homogenate-samples** performed with transcripts (n=24) which were detected in MECs+GECs of almost (six of seven) samples (Chapter 6.4.4), excluding the transcripts of CYP1A1, CYP19A1, UGT1A8, UGT2B7, UGT1A3/4, SULT1A1, SULT2A1, SULT1A2, SULT1E1, HSD17B1, ESR2, NR1I3, MKI67, BMF and NR1I2. Transcript level were standardized prior to cluster analysis (Function "scale()" in the software R, Chapter 5.2.15).

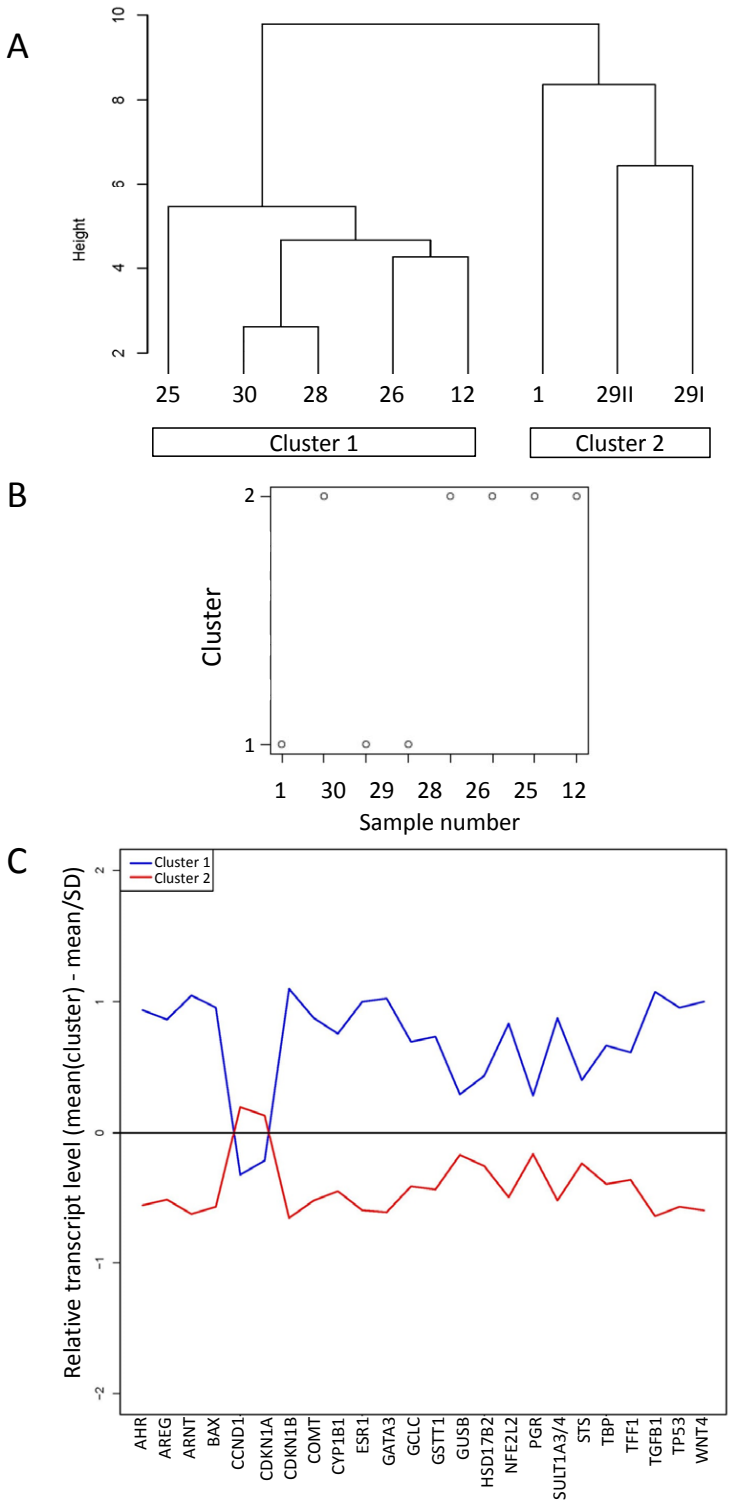


Figure 75: Dendrogram (A), cluster classification (B) and characteristics of the two main clusters concerning their relative (n0/n0 HPRT) transcript levels presented as mean(cluster)-mean(all)/SD(all) (C) of the cluster analysis of transcripts levels of all eight **MECs+GECs-samples** performed with transcripts (n=24) which were detected in MECs+GECs of almost (six of seven) samples (Chapter 6.4.4), excluding the transcripts of CYP1A1, CYP19A1, UGT1A8, UGT2B7, UGT1A3/4, SULT1A1, SULT2A1, SULT1A2, SULT1E1, HSD17B1, ESR2, NR1I3, MKI67, BMF and NR1I2. Transcript levels were standardized prior to cluster analysis (Function "scale()" in the software R, Chapter 5.2.15).

9.6 Metabolic Network

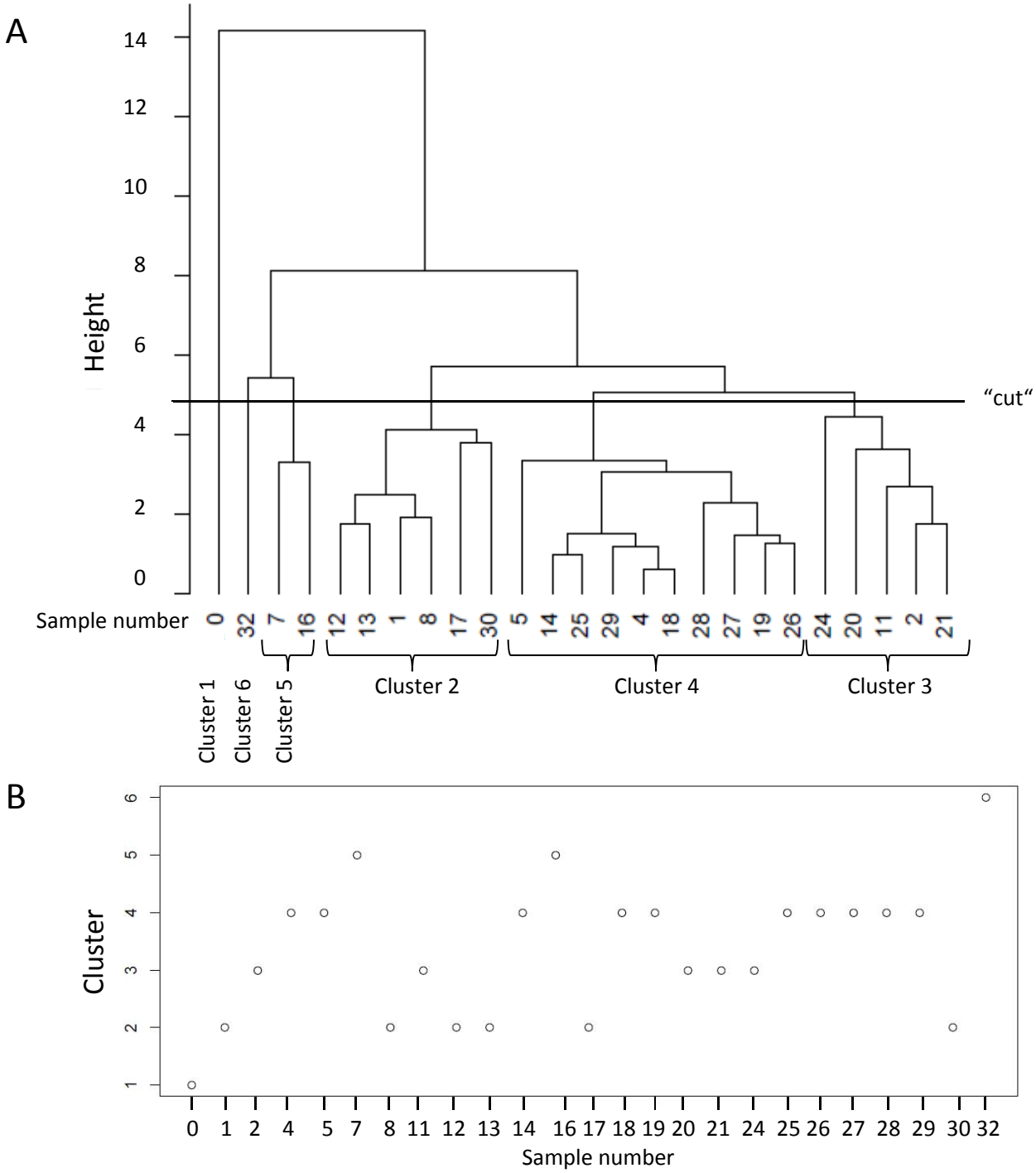


Figure 76: Cluster analysis of 25 samples with calculated fluxes of CYP1A1, CYP1B1, SULT1A1, SULT1A2, SULT1E1, SULT2A1, COMT, GSTP1, NQO1, HSD17B1, HSD17B2, STS, GCLC and E1/E2 DNA adduct formation fluxes. A: Dendrogramm. B: Cluster classification of the samples.

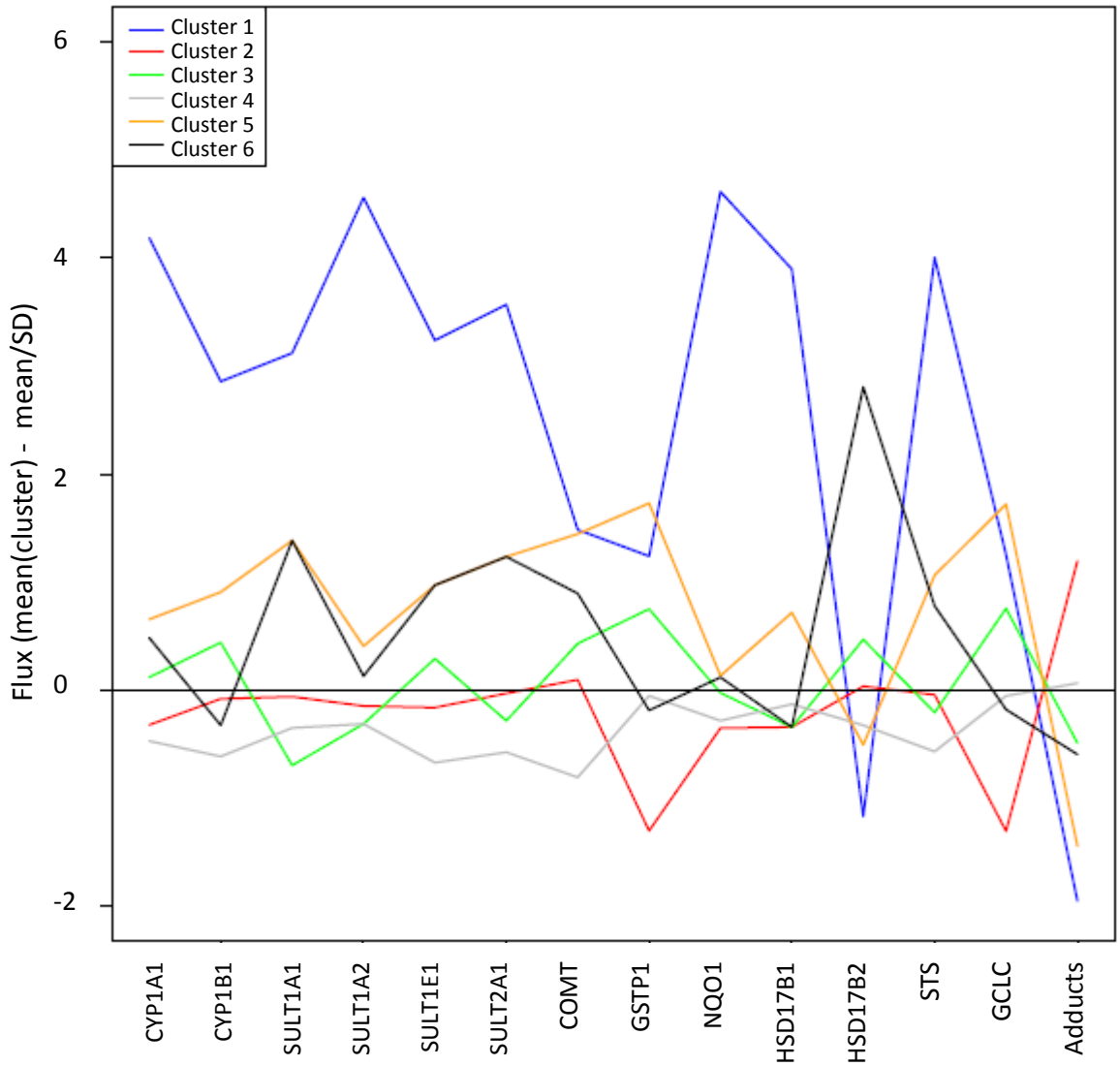


Figure 77: Characteristics of all clusters concerning their calculated flux values of enzymes and DNA adducts presented as mean(cluster)-mean(all)/SD(all).

Table 67: DNA adduct forming fluxes with (+ pol) and without (- pol) consideration of polymorphisms for network calculation.

Sample #	DNA adduct forming flux - pol	+ pol	Increase (%)
0	<0.0004	<0.0004	-
1	0.0276	0.0276	100.00
4	0.0292	0.0292	100.00
5	0.0164	0.0164	100.00
11	0.0212	0.0212	100.00
12	0.0352	0.0352	100.00
13	0.0284	0.0288	101.41
14	0.0240	0.0248	103.33
16	0.0004	0.0004	100.00
17	0.0360	0.0368	102.22
18	0.0296	0.0296	100.00
21	0.0140	0.0144	102.86
24	0.0092	0.0168	182.61
25	0.0240	0.0240	100.00
26	0.0164	0.0164	100.00
27	0.0148	0.0168	113.51
28	0.0228	0.0228	100.00
29	0.0280	0.0280	100.00
30	0.0360	0.0360	100.00
32	0.0144	0.0144	100.00

9.7 Statistic mRNA level breast tissue samples

Table 68: Correlation coefficients of correlation analysis of transcripts, age, BMI, E2 and E1 tissue levels, sum of E1 and E2 and ratio E2/E1 with R>0.5.

Transcript/factor	R	Transcript/factor
Age	0.57	CDKN1A
Age	0.57	WNT5A
AHR	0.60	BAX
AHR	0.57	CDKN1B
AHR	0.58	TGFB1
ARNT	0.54	BAX
ARNT	0.66	BMF
ARNT	0.74	CDKN1B
ARNT	0.66	GCLC
ARNT	0.71	TGFB1
BAX	0.78	CDKN1B
BAX	0.65	TGFB1
BMF	0.69	CDKN1B
BMF	0.68	TGFB1
BMI	-0.52	ARNT
BMI	-0.71	GSTP1
BMI	-0.56	HSD17B1
CCND1	0.55	CDKN1B
CCND1	0.61	STS
CDKN1A	0.63	ARNT
CDKN1A	0.62	CDKN1B
CDKN1A	0.64	ESR2
CDKN1A	0.72	GCLC
CDKN1A	0.53	HSD17B1
CDKN1A	0.73	NFE2L2
CDKN1A	0.53	SULT1A3.4
CDKN1A	0.60	TGFB1
COMT	0.51	GCLC
CYP1A1	0.55	NQO1
E2	0.67	E1
E2	0.53	PGR
E2	0.89	Sum_E1_E2
ESR2	0.55	ARNT
ESR2	0.62	BMF
ESR2	0.59	CDKN1B
ESR2	0.76	GCLC
ESR2	0.54	SULT1A2
ESR2	0.51	TGFB1
ESR1	0.53	Age
ESR1	0.57	AHR
ESR1	0.53	ARNT
ESR1	0.51	CDKN1A
ESR1	0.58	GATA3
ESR1	0.64	TGFB1
GATA3	0.51	ARNT
GATA3	0.73	CDKN1A
GATA3	0.51	GCLC
GATA3	0.58	TGFB1
GATA3	0.61	WNT5A
GCLC	0.51	BAX
GCLC	0.54	BMF
GCLC	0.70	CDKN1B
GCLC	0.62	TGFB1
GSTT1	0.67	CDKN1A
GSTT1	0.51	CDKN1B
GSTT1	0.54	ESR2
GSTT1	0.58	GCLC
GSTT1	0.67	NFE2L2
HSD17B1	0.63	ARNT
HSD17B1	0.83	ESR2
HSD17B1	0.66	GCLC
MKI67	0.63	BMI
MKI67	-0.54	ESR1
MKI67	-0.69	GSTP1
MKI67	-0.50	NQO1
MKI67	0.61	TP53
MKI67	0.58	WNT4
NFE2L2	0.51	ARNT
NFE2L2	0.53	BMF
NFE2L2	0.72	CDKN1B
NFE2L2	0.69	ESR2
NFE2L2	0.67	GCLC
NFE2L2	0.62	STS
NFE2L2	0.50	SULT1A2
NFE2L2	0.50	SULT1A3.4
NFE2L2	0.56	TGFB1
NQO1	0.54	ESR2
NQO1	0.57	GCLC
NQO1	0.46	HSD17B1
NQO1	0.52	SULT1A1
NQO1	0.60	UGT2B7
NR1I2	-0.63	Age
NR1I2	-0.50	GSTP1
Ratio_E2_E1	0.66	E2
STS	0.52	BAX
STS	0.64	CDKN1B
SULT1A1	0.52	ARNT
SULT1A1	0.66	CDKN1A
SULT1A1	0.60	ESR2
SULT1A1	0.56	GCLC
SULT1A1	0.54	HSD17B1
SULT1A1	0.59	NFE2L2
SULT1A1	0.53	SULT1A2
SULT1A2	0.53	ARNT
SULT1A2	0.57	CDKN1B
SULT1A2	0.60	GCLC
SULT1A2	0.61	TGFB1
SULT1A3.4	0.50	AHR
SULT1A3.4	0.67	BMF
SULT1A3.4	0.52	CDKN1B
SULT1E1	0.71	Age
SULT1E1	0.58	WNT5A
Sum_E1_E2	0.91	E1
Sum_E1_E2	-0.55	ESR1
Sum_E1_E2	0.53	PGR
TFF1	0.80	AREG
TFF1	-0.51	ESR2
TGFB1	0.83	CDKN1B
TP53	0.56	AREG
TP53	0.50	NR1I2
UGT1A8	0.50	BMF
UGT2B7	0.54	WNT5A
WNT4	-0.51	Age
WNT4	0.61	AREG
WNT4	-0.62	ESR1
WNT4	0.53	TFF1
WNT4	0.60	TP53
WNT5A	0.70	CDKN1A
WNT5A	0.54	GSTT1
WNT5A	0.51	NFE2L2

Table 69: Two-sample comparison p values and adjusted p values with method of "Holm" for comparison of transcript levels between smokers and non-smokers (_S) and lobule type 1_{parous} and other lobule types.

Transcript	p values	adjusted p values
COMT_S	0.8795	1
COMT_L	0.8784	1
WNT5A_S	0.4099	1
WNT5A_L	0.009164	0.742284
WNT4_S	0.8795	1
WNT4_L	0.1061	1
UGT2B7_S	0.4321	1
UGT2B7_L	0.02369	1
SULT1A2_S	0.8747	1
SULT1A2_L	0.06546	1
SULT1E1_S	0.3193	1
SULT1E1_L	0.01716	1
AHR_S	0.119	1
AHR_L	0.8064	1
ARNT_S	0.1584	1
ARNT_L	0.6722	1
CCND1_S	0.5303	1
CCND1_L	0.3971	1
CDKN1A_S	0.6486	1
CDKN1A_L	0.005078	0.416396
CYP1A1_S	0.1196	1
CYP1A1_L	0.3684	1
ESR1_S	0.3193	1
ESR1_L	0.01372	1
GSTP_S	9.83E-05	0.0082572
GSTP_L	0.04138	1
HSD17B1_S	0.08491	1
HSD17B1_L	0.3684	1
MKI67_S	0.002038	0.169154
MKI67_L	0.04304	1
NR1I2_S	0.7742	1
NR1I2_L	0.01671	1
NR1I3_S	0.1404	1
NR1I3_L	0.2666	1
PGR_S	0.7616	1
PGR_L	0.1417	1
SULT1A1_S	0.4929	1
SULT1A1_L	0.01086	0.8688
TP53_S	0.8124	1
TP53_L	0.02908	1
TFF1_S	0.7043	1
TFF1_L	0.1229	1
BMF_S	0.8795	1
BMF_L	0.7309	1
BAX_S	0.5426	1
BAX_L	0.6796	1
GCLC_S	0.5946	1
GCLC_L	0.02618	1
TGFB1_S	0.2827	1
TGFB1_L	0.2982	1
AREG_S	0.4008	1
AREG_L	0.2372	1
GATA3_S	0.8201	1
GATA3_L	0.1229	1
CDKN1B_S	0.4456	1
CDKN1B_L	0.4475	1
NFE2L2_S	0.3027	1
NFE2L2_L	0.02858	1
ESR2_S	1	1
ESR2_L	0.03868	1
STS_S	0.9396	1
STS_L	0.8907	1
HSD17B2_S	0.8795	1
HSD17B2_L	0.4902	1
GSTT1_S	0.3695	1
GSTT1_L	0.7897	1
NQO1_S	0.2033	1
NQO1_L	0.441	1
SULT1A3/4_S	0.7903	1
SULT1A3/4_L	0.4377	1
UGT1A10_S	0.905	1
UGT1A10_L	0.1195	1
UGT1A7_S	0.8791	1
UGT1A7_L	0.696	1
UGT1A6_S	1	1
UGT1A6_L	0.02535	1
UGT1A1_S	0.2232	1
UGT1A1_L	0.3059	1
UGT1A8_S	0.4554	1
UGT1A8_L	0.6459	1
CYP19A1_S	0.6516	1
CYP19A1_L	0.4078	1
CYP1B1_S	0.9924	1
CYP1B1_L	0.3184	1

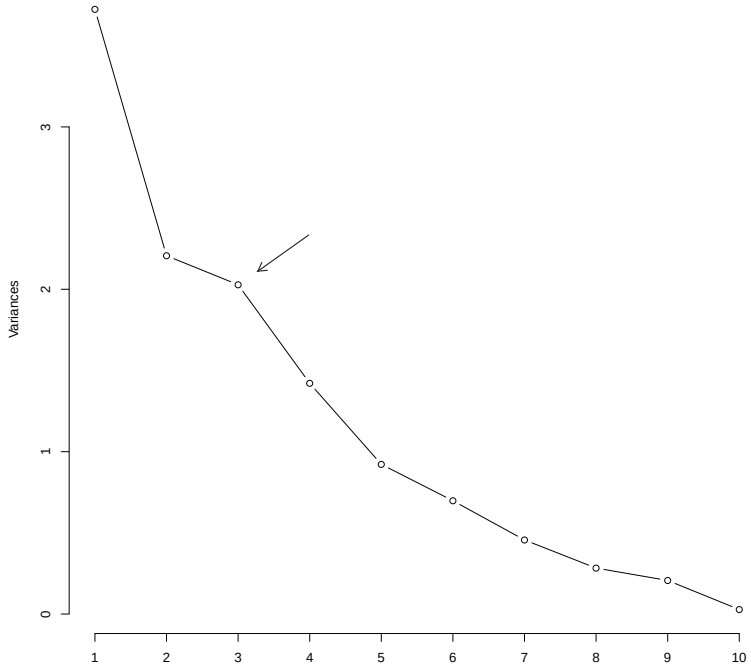


Figure 78: Scree plot of PCA of transcripts of E2 metabolizing enzymes. The "Elbow point" (arrow) is at 3 principal components. X-axes, number of principal components. Transcript levels used for PCA: CYP1A1, CYP1B1, UGT1A1, UGT1A8, UGT2B7, SULT1A1, SULT1A2, SULT1E1, STS, CYP19A1, HSD17B1, HSD17B2.

Further supplementary data can be found in the accompanying data disc.